Signalling and Prediction of Failures in Discrete Control Devices with Structural Redundancy

M. A. GAVRILOV

In solving problems of providing reliable operation of automatic control devices, a great deal of attention is devoted to the use of methods involving the application of structural redundancy. These include all possible methods of duplicating individual elements within units, as well as the more common methods of providing redundancy of all the necessary elements and units on the whole with the least possible number of additional elements. The ever-increasing practical use of methods of structural redundancy is a result of the fact that, in present complex automatic systems, the control devices require such a large number of individual elements to perform their functions that even though the elements may have a very high reliability, the necessary reliability demanded of the entire device cannot be achieved.

A number of works[1–6] is devoted to the question of the introduction of structural redundancy and the determination of the minimum number of additional elements necessary to achieve the prescribed reliability of the device on the whole. For discrete control devices it is most natural and suitable to examine the required value of operating reliability of the device as being prescribed by a certain number of elements which simultaneously fail during operation while nevertheless permitting the device to perform accurately the control algorithm assigned to it[7].

The author of the present report showed[3] that when the problem is treated in this manner, the determination of the minimum number of additional internal elements necessary to achieve a given reliability completely coincides with the task of determining the minimum number of additional symbols in the construction of correcting codes with correction of the corresponding number of errors. In the same article a method was given for constructing tables of states which provide for a realization of the structure of a discrete control device having the required reliability.

The proposed method links the problem of constructing such a device to the distrubution of the states of its internal elements along the vertices of a many-dimensional cube of single transitions in such a manner that the number of transitions (distance) between the vertices, selected for the corresponding stable states of the device, would be no less than:

$$D=2d+1 \qquad (1)$$

where d is the number of simultaneously failing elements with which the devices must still exactly perform their control algorithm.

In differentiating the demands on reliability (namely, separating them from the viewpoint of the number of simultaneously failing elements), first, into that for which the device must accurately perform a given control algorithm and, second, into that for which it must not provide any actions at its outputs, the value of the distance between vertices selected for the stable states must be no less than:

$$D=2d+\Delta+1 \qquad (2)$$

where Δ is the number of simultaneously failing elements in addition to d for which the indicated second condition of reliable operation of the device must be fulfilled.

In discrete types of devices which have reliability as a result of structural redundancy, the required reliability is retained only until the moment of onset of permanent failure of even one of the elements.

In fact, let the prescribed probability of failure of the entire device on the whole require that the given control algorithm be exactly performed with the simultaneous failure of d elements. Then, with a permanent failure of any one of the elements, the device will capably perform the control algorithm only upon the simultaneous failure of $d-1$ elements; that is, it will have a probability of complete failure which is less than prescribed.

Particular importance is therefore devoted to rapid signalling of failure of individual elements or their prediction, which permits one to take timely measures to replace the faulty elements or other measures which will return the probability of failure of the entire device to its prescribed value. The present report is devoted to an examination of the fundamental possibilities of providing such signalling or prediction for automatic control devices designed on the basis of the principles described by the author[3].

First it is shown that the table of states constructed according to the principles contains all the necessary information on failure, both generally for all elements as well as for each of them individually, and, even more, on the nature of the failures.

Those states of internal elements which correspond to the stable states of the corresponding table of transitions and which are distributed, as was pointed out above, in the vertices of a many-dimensional cube of single transitions with a distance one from the other of not less than D, are called basic. To each of these states there must correspond a particular state of outputs which provides for the performance of the prescribed control algorithm.

Let the number of inputs of the discrete device be equal to a and let it be given that, to perform the control algorithm with a prescribed degree of reliability, that is, in the presence of simultaneous failure of d internal elements, it is necessary to have K internal elements. Then each of the basic states may be characterized by a certain conjunctive member of a Boolean function of length $a+K$. In accordance with this the table of states contains, on the left-hand side, $a+K$ columns of

which a characterizes the states of the inputs and K characterizes the states of the internal elements. The binary number characterizing the state of the internal elements corresponds to a particular vertex of the many-dimensional cube, selected in distributing the given basic state.*

The failure of any element is characterized by a change in the binary number, corresponding to a given basic state, from zero to one or one to zero. The first is called a $0 \rightarrow 1$ type failure and the second a $1 \rightarrow 0$ type failure. Each such failure transfers the basic state to an adjacent vertex of the many-dimensional cube. The simultaneous failure of any two internal elements transfers the basic state to a vertex two units removed from the vertex selected for the given basic state; it is adjacent to any vertex to which the basic state was transferred by the failure of any one of these two elements.

In order to provide exact performance of the control algorithm upon the failure of internal elements, each of the states to which the basic state is transferred upon the failure of any number of elements within the prescribed limits (that is, inclusive to d) must compare in the right-hand side of the table of states to the same state of outputs as the basic state. Therefore, for each stable state of the table of transitions, for the case of structural redundancy, there must correspond a particular combination of states consisting of the basic state and all the states to which it transfers upon failure of the internal elements. All of these states are adjacent to one another, forming a certain multiple of adjacent states. This multiple is called a set of basic states.

Frist it is shown that the set of adjacent states, together with the basic states, may be described by a symmetrical Boolean function whose active numbers represent a natural series of numbers from $K - d$ to K.

Let there be any state f_{i0} corresponding to one of the basic states and let this state be characterized by a row in the table of states containing K_1 zeros and K_2 ones, where $K_1 + K_2 = K$. Then, with $d = 1$, the collection of adjacent states Σf_{i1} contains all the states differing from the basic by the replacement of one variable by its reciprocal. More precisely, they are K, while K_1 of them corresponds to a failure of the type $0 \rightarrow 1$ and K_2 to a failure of the type $1 \rightarrow 0$. It is easy to see that the sum of these states may be characterized by the symmetrical function:

$$\Sigma f_{i1} = S_{K-1}(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{K_1}, \bar{x}_{K_1+1}, x_{K_1+2}, \ldots, x_{K_1+K_2})$$

if the basic state is considered a symmmetrical function of those variables with an active number equal to K, namely:

$$f_{i0} = S_K(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{K_1}, x_{K_1+1}, x_{K_1+2}, \ldots, x_{K_1+K_2})$$

The sum of the basic and set of adjacent states is thus characterized by the symmetrical Boolean function:

$$f_{i0} + \Sigma f_{i1} = S_{K-1,K}(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_K, x_{K+1}, x_{K+2}, \ldots, x_{K_1+K_2})$$

If $d = 2$, the set of adjacent states consists of all states differing from the basic by the replacement of one variable by its reciprocal, the number of which, as was pointed out, is equal to $K = C^1_K$, and two variables. The number of the latter is obviously equal C^2_K, and since each of them differs from the basic by a change having a value of two variables, their total Σf_{i2} corresponds to the symmetrical function:

$$\Sigma f_{i2} = S_{x-2}(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{K_1}, \bar{x}_{K_1+1}, x_{K_1+2}, \ldots, x_{K_1+K_2})$$

The Boolean function characterizing the basic state and the entire set of adjacent states is thus a symmetrical function of the type:

$$f_{i0} + \Sigma f_{i1} + \Sigma f_{i2} = S_{K-2,K-1,K}(\bar{x}, \bar{x}_2, \ldots, \bar{x}_{K_1}, x_{K_1+1}, x_{K_1+2}, \ldots, x_{K_1+K_2})$$

It may be proved in an analogous manner that in the general case, with the simultaneous failure of d internal elements, the basic state and the set of adjacent states may be characterized by a symmetrical Boolean function of the type:

$$S_{K-d,K-d+1,\ldots,K}(\bar{x}_1, \bar{x}_2, \ldots, x_{K_1}, x_{K_1+1}, x_{K_1+2}, \ldots, x_{K_1+K_2})$$

Thus, the class of reliable structures of discrete devices is, with respect to internal elements, a class described by symmetrical Boolean functions of a special type, which facilitates their realization since these functions have been most widely studied and may be economically realized with the aid of different types of threshold relay elements, including electromagnetic relay elements with several windings[8].

The basic state is designated as f_i and the set of adjacent states corresponding to it as N_i, assuming that $f_i + N_i = F_i$.

The table of states of a discrete control device consists on the left-hand side of all sets F_i combined with the corresponding values of inputs. For each of these sets there corresponds on the right-hand side of the table, as was pointed out above, a state of outputs which provides for the performance of the control algorithm. One more output is added for which is included in the table of states a zero for each of the basic states and a one for any of the states which are included in the sets of adjacent states.

Since the latter corresponds to the failure of any one or to the simultaneous failure of several internal elements, the appearance of a one at this output occurs only by means of a decrease in the reliability of operation of the discrete device and may be used to signal the presence of a failure.

For example, let there be a discrete device with three inputs and one output (*Figure 1*) and an action, equal to one, must appear at the latter in the subsequent sequence of change of the states of the outputs:

0 0 0
1 0 0
1 1 0
1 1 1
0 1 1

Any subsequent change of inputs must lead to the appearance of an action at an output to zero, while the further appearance of an action at the output equal to one occurs only by the repetition of the indicated sequence of change of the states of the inputs. With any other sequence of change of the states of the inputs, the action at the output must remain equal to zero.

The corresponding table of conversions is given in *Table 1*. Here it may be seen that it is necessary to provide for four stable states, which is possible with the aid of two internal elements.

When it is necessary that the aforementioned discrete device performs exactly a preassigned control algorithm in the event of the simultaneous failure of one of the internal elements, five

* All references made below to internal elements with an identical base pertain to inputs and sensing elements.

Table 1

000	100	110	010	011	111	101	001
$(1)^0$	$(1)^0$	2	4	$(1)^1$	4	4	4
—	4	$(2)^0$	4	—	3	—	—
—	—	4	—	1	$(3)^0$	4	—
1	$(4)^0$	$(4)^0$	$(4)^0$	$(4)^0$	$(4)^0$	$(4)^0$	$(4)^0$

internal elements are required, as seen in *Table 5* of reference 3.

The following distribution for the basic states is chosen:

0 0 0 0 0
1 0 1 1 0
0 1 0 1 1
1 1 1 0 1

Then the table of states will have the form shown in *Table 2*. In agreement with what was mentioned above, let us add the output C_0, in the column of which are written zeros in all the rows of the table of states corresponding to f_i and ones in all the rows corresponding to N_i (*Table 3*). Then this output will signal the presence of a failure of any one or several of the internal elements.

Table 2

A	B	C	F	X_1	X_2	X_3	X_4	X_5	Z
0	0	0	F_2	0	0	0	0	0	0
0	0	0	F_4	0	0	0	0	0	0
0	0	1	F_1	1	1	1	0	1	0
0	0	1	F_4	1	1	1	0	1	0
0	1	0	F_1	1	1	1	0	1	0
0	1	0	F_2	1	1	1	0	1	0
0	1	0	F_4	1	1	1	0	1	0
0	1	1	F_1	0	0	0	0	0	1
0	1	1	F_3	0	0	0	0	0	1
0	1	1	F_4	1	1	1	0	1	0
1	0	0	F_1	0	0	0	0	0	0
1	0	0	F_2	1	1	1	0	1	0
1	0	0	F_4	1	1	1	0	1	0
1	0	1	F_1	1	1	1	0	1	0
1	0	1	F_3	1	1	1	0	1	0
1	0	1	F_4	1	1	1	0	1	0
1	1	0	F_1	1	0	1	1	0	0
1	1	0	F_2	1	0	1	1	0	0
1	1	0	F_3	1	1	1	0	1	0
1	1	0	F_4	1	1	1	0	1	0
1	1	1	F_1	1	1	1	0	1	0
1	1	1	F_3	0	1	0	1	1	0
1	1	1	F_2	0	1	0	1	1	0
1	1	1	F_4	1	1	1	0	1	0

In this table:

F_1	F_2	F_3	F_4
0 0 0 0 0	1 0 1 1 0	0 1 0 1 1	1 1 1 0 1
1 0 0 0 0	0 0 1 1 0	1 1 0 1 1	0 1 1 0 1
0 1 0 0 0	1 1 1 1 0	0 0 0 1 1	1 0 1 0 1
0 0 1 0 0	1 0 0 1 0	0 1 1 1 1	1 1 0 0 1
0 0 0 1 0	1 0 1 0 0	0 1 0 0 1	1 1 1 1 1
0 0 0 0 1	1 0 1 1 1	0 1 0 1 0	1 1 1 0 0

Table 3

X_1	X_2	X_3	X_4	X_5	C_0	C_1	C_2	C_3	C_4	C_5
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	1	0	0	0	0
0	1	0	0	0	1	0	1	0	0	0
0	0	1	0	0	1	0	0	1	0	0
0	0	0	1	0	1	0	0	0	1	0
0	0	0	0	1	1	0	0	0	0	1
1	0	1	1	0	0	0	0	0	0	0
0	0	1	1	0	1	1	0	0	0	0
1	1	1	1	0	1	0	1	0	0	0
1	0	0	1	0	1	0	0	1	0	0
1	0	1	0	0	1	0	0	0	1	0
1	0	1	1	1	1	0	0	0	0	1
0	1	0	1	1	0	0	0	0	0	0
1	1	0	1	1	1	1	0	0	0	0
0	0	0	1	1	1	0	1	0	0	0
0	1	1	1	1	1	0	0	1	0	0
0	1	0	0	1	1	0	0	0	1	0
0	1	0	1	0	1	0	0	0	0	1
1	1	1	0	1	0	0	0	0	0	0
0	1	1	0	1	1	1	0	0	0	0
1	0	1	0	1	1	0	1	0	0	0
1	1	0	0	1	1	0	0	1	0	0
1	1	1	1	1	1	0	0	0	1	0
1	1	1	0	0	1	0	0	0	0	1

If one places the action from this output into a computer and determines the number of times that actions equal to one appear at this output during a certain time interval, the answers from the computer may be used to predict an approximation of reliable operation of the device.

The described principle of signalling and prediction has significant advantages in the sense that neither the signalling nor prediction requires the introduction of any additional internal elements. Usually the performance of these functions relies upon special units of the discrete device which require elements having, in principle, a reliability as much as one order of magnitude greater than the elements which make up the discrete device itself.

In the design examined above, comprising a structure of signal outputs based on actuating devices already having internal elements, and assuming that the connections between these devices and the sensing signal and predicting devices have 100 per cent reliability, one would expect that the signalling of failure would have absolute reliability in principle.

In fact, only two mutually exclusive events may occur: (*a*) not one of the internal elements is faulty. Then the actions equal to one appear at the corresponding operating outputs and at the signal output the action is equal to zero; (*b*) failure of one or several internal elements occurs within the limits of *d*. Then an action equal to one appears both at the signal and operating outputs.

It is noted that achieving reliable operation by means of the introduction of structural redundancy according to the principles previously presented by the author[3] pertain to the internal elements of the device as a whole, that is, both to the actuating

and the reacting devices. Therefore, with respect to failures of the actuating organs, the device retains its ability to perform exactly the control algorithm upon the failure of either one or, simultaneously, all of the actuating devices of a given internal element for the conditions when these failures are all of a single type.

The described principle of designing signal circuits makes it possible to provide separately for signalling the number of failures greater than d, including those located between the limits of $d+1$ and $d+\Delta$. Additional outputs must be added for this purpose. This requires that ones be written in the specific rows in the appropriate columns of the table of states; namely, for signalling failures of elements within limits from $d+1$ to $d+\Delta$ in the rows corresponding to failures in these limits, and for signalling a large number of failures in the rows corresponding to unused states.

It is obvious that the signalling of failures may be not only general but also specific, or, for each of the internal elements of the device separately. For this purpose one must have for each of them an individual output, for which there must be written in the columns of the table of states *ones* for all states differing from the basic by the change in value of the corresponding variable. For example, to signal the failure of element X_1 in the above case, ones must be written for each first row of the sets N_i for the corresponding output.

Table 3 gives the corresponding values of outputs for each of the internal elements. The realization of such outputs provides, in the event of faulty elements in the device, for advance notification as to which of the internal elements is malfunctioning or, with prediction, an approximate indication, permitting timely replacement or adjustment of the element for proper action.

Obviously it is possible to provide not only for signalling of failures of individual internal elements but for the separate signalling of the nature of these failures as well. For example, in *Table 4*, for the internal element X_1 examined above, are shown the operating states corresponding to failures of the type $0\rightarrow 1$ [*Table 4(a)*] and failures of the type $1\rightarrow 0$ [*Table 4(b)*].

Table 4

X_1	X_2	X_3	X_4	X_5
1	0	0	0	0
1	1	0	1	1

(*a*)

X_1	X_2	X_3	X_4	X_5
0	0	1	1	0
0	1	1	0	1

(*b*)

In conclusion some of the problems of realizing signalling and prediction networks are considered. The circuit of each output in the structure of a multi-cycle discrete device must contain actuating devices of both internal and sensing relay elements. The signal circuits must contain actuating devices of only internal elements. Therefore the rational design of the structure of a discrete device would be that shown in *Figure 2*, namely, a structure in the form of a certain $[1,K]$ terminal network having at its outputs all the functions of f_i and N_i and containing the actuating devices of only the internal elements, and an $[M,N]$ terminal network containing the actuating devices of only the sensing elements.

As was pointed out above, the functions which realize the basic states together with the sets of adjacent states are symmetrical with the operating numbers from $K-d$ to K and for their realization it is suitable to use so-called 'threshold' elements. When such elements are used it is advantageous to use the structure of the discrete device having a form shown in *Figure 2(b)*, where the $[1,K]$ terminal network is based on threshold elements according to the number of basic states. The $[M,N]$ terminal network has the same make-up as that shown in *Figure 2(a)*, while the output circuits for signalling and prediction of failures are derived from the outputs of the threshold elements by means of their series connection (providing an 'and' operation) and from circuits corresponding to the function $\bar{f}_i$. The latter may also be designed with the aid of threshold elements having symmetrical functions with the operating number K.

In addition it is noted that, in the case examined above, it is most rational from the viewpoint of the simplest physical realization of the structure of a discrete device to choose the operating levels of the symmetrical functions not from $K-d$ to K but from 0 to d, while simultaneously taking not the variables but their inversions.

In conclusion one should note that the method considered previously by the author[3], as well as everything discussed in this report, refer to the case in which the probability of failure for all internal elements has a single value, the failures are symmetrical (that is, the probability of failures of the type $0\rightarrow 1$ is identical to that of type $1\rightarrow 0$), and, in addition, failures of individual elements are mutually independent. Conditions differing from these necessitate a somewhat different approach to determining the minimum number of elements and the distribution of the states. However, the principles of designing signal circuit and of prediction remain the same, with the exception that the functions characterizing the basic sets and the sets of adjacent states may not prove symmetrical.

References

[1] VON NEUMAN, S. Probabilistic logics and the synthesis of reliable organisms from unreliable components. *Automata Studies*. 1956. Princeton; Princeton University Press

[2] MOORE, E. F. and SHANNON, C. E. Reliable circuits using less reliable relays. *J. Franklin Inst.* Vol. 262, No. 3 (1956) 191, 281

[3] GAVRILOV, M. A. Structural redundancy and reliability of relay circuits. *Automatic and Remote Control.* Vol. 2, p. 838. 1961. London; Butterworths

[4] ZAKROVSKIY, A. D. A method of synthesis of functionally stable automata. *Dok. AN SSSR* Vol. 129, No. 4 (1959) 729

[5] RAY-CHANDHURI, D. K. On the construction of minimally redundant reliable system designs. *B.S.T.J.* Vol. 40, No. 2 (1961) 595

[6] ARMSTRONG, D. B. A general method of applying error correction to synchronous digital systems. *B.S.T.J.* Vol. 40, No. 2 (1961) 577

[7] GAVRILOV, M. A. Basic terminology of automatic control. *Automatic and Remote Control.* Vol. 2, p. 1052. 1961. London; Butterworths

[8] GAVRILOV, M. A. *The Structural Theory of Relay Devices, Part 3. Contactless Relay Devices.* 1961. Moscow; Publishing House of the All Union Correspondence Power Engineering Institute

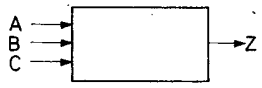
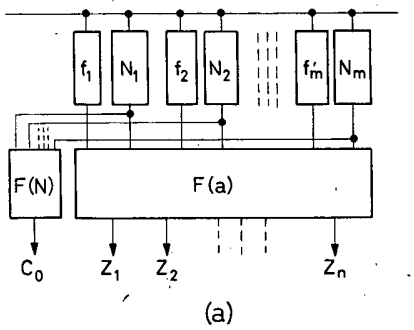


Figure 1



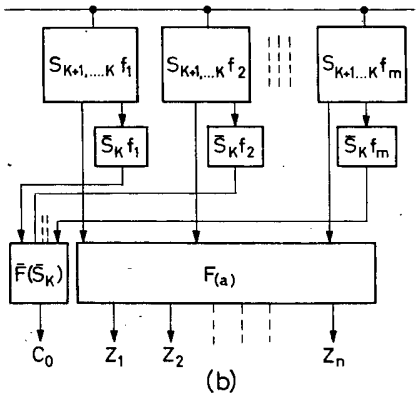


Figure 2

Dup-

A Digital Optimal System of Programmed Control and its Application to the Screw-down Mechanism of a Blooming Mill

S. M. DOMANITSKY, V. V. IMEDADZE and Sh. A. TSINTSADZE

Introduction

Digital servo programmed-control systems are finding continually wider applications in various branches of industry: in particular, they are used for the automatic control of screw-down and other mechanisms of rolling mills, for the control of various moving parts in control systems for metal-cutting machine tools, and in a number of other instances. The operation of such mechanisms normally falls into two stages. In the first stage the device must choose or compute an optimal programme, working on the basis of information about the requirements for the technological process, about the condition of the plant, about external perturbations, etc. In the second stage the given programme must be carried out according to an optimal law. The term 'optimal law' is normally taken to mean the carrying out of the given displacements with the maximum possible response speed and with the required accuracy; in addition a condition is often included covering requirements on control response quality.

While the function of choosing an optimal programme is not necessarily inherent in the digital control system itself, particularly when it operates in a complex installation with a controlling computer, the function of carrying out the given displacements according to an optimal law must still be organically inherent in the digital servo system. If this requirement is not satisfied, such systems cannot be considered fully efficient, since for many mechanisms, e.g. manipulator jaws, shears and rolling-mill pressure screw-down, the response speed and accuracy determine the productivity and output quality of the whole line.

A system of programmed control has been developed by the Institute of Electronics, Automatic and Remote Control of the Academy of Sciences of the Georgian S.S.R. in cooperation with the Institute of Automatic and Remote Control of the U.S.S.R. Academy of Sciences. The basic unit of this system is a digital optimal servo system which has a number of characteristic properties. The electric motor drive of the optimal system works at accelerations that are maximal and constant in magnitude. This ensures the greatest response speed and simplifies the design of the computing part of the programmed-control system. The required system accuracy is ensured by the digital form in which the programme is given and executed. The small quantity of information processed in unit time has made it possible to use a pulse-counting code rather than a binary one, which improves the reliability and interference-rejection properties of the system. The system is entirely built out of ferrite and transistor elements.

This report gives a general description of the digital optimal programmed-control system, and also a practical example of its application to the automatic control of the screw-down mechanism of a blooming mill; this device has passed through laboratory and factory testing, and by the end of 1962 it was introduced into experimental service at the Rustavi steelworks.

Design Principles of the Programmed-control System

The basis of the system developed for programmed control is the optimum principle; the execution of the required displacement takes place at limiting values of the restricted coordinates, especially of the torque and rotation speed of the motor.

For the case where the drive control system has negligible inertia, *Figure 1* will clarify the above; it shows the law taken for the variation of the control action F_y, and the curves of motor torque M_m and speed n. The figure shows that during run-up and braking the drive maintains the constant maximum permissible value of torque developed by the motor. When executing large displacements, after the motor has reached its maximum speed n_{max} it is automatically switched over by the drive circuit to operate at that constant speed (point *MS* on *Figure 1*). The instant of braking (point T_2) is chosen by the control system such that only a relatively short path remains to be traversed up to the instant when the speed is reduced to 10–12 per cent of the maximum (point *CS*). The execution of the rest of the path to the required low speed is automatically performed by the drive circuit, and ensures maximum accuracy in carrying out the programme. *Figure 1* shows that the variation of drive speed with time follows a trapezoidal law. For small required displacements the motor does not have time to run up to its maximum speed, and the speed variation follows a triangular law.

The above-mentioned properties of the drive allow the controlling part of the programmed-control system to be considerably simplified, since in this event it only has to generate and execute commands for starting the drive in the required sense, for braking and for stopping the drive.

The design logic is very simple for that part of the control system whose purpose is to start the drive in the required sense and to determine the instant for generating the command to stop the drive; it is suitable both for control of low-power drives that have no links with appreciable inertia, and also for control of high-power drives with large inertia. The required displacement path and sense of rotation of the motor are determined by comparing the given programme with the actual position of the controlled mechanism (to give an error signal). During the execution process the path traversed is continuously compared with the initial error; the command to stop the drive is generated at the instant when these two quantities become equal.

The programme is given in terms not of previously defined initial errors, but of absolute values of position-coordinates for the controlled mechanism. This avoids the possibility of errors accumulating from excution to execution, and also the need for the controlled mechanism to be resting initially in a closely defined position.

The part of the system that determines the instant for the command to start braking has a relatively more complex design logic, and also takes different forms in systems for controlling the two different types of drive mentioned above.

For systems controlling inertia-free drives the ratio between the paths traversed on braking S_b and on the run-up S_r is a constant and equal to the ratio between the absolute values of acceleration on run-up a_r and on braking a_b:

$$\frac{S_b}{S_r}=\frac{a_r}{a_b}=k \tag{1}$$

Taking into account the condition that should be satisfied:

$$S_r+S_b=\Delta$$

where Δ is the required excution path (i.e. the initial error), one gets

$$\Delta=S_r(1+k) \tag{2}$$

This expression defines the design logic for the part of the system determining the instant for the command to start braking: the path traversed by the drive during the run-up is continuously multiplied by the fixed quantity $1+k$, and when the resultant quantity becomes equal to the initial error Δ, then the command is generated to start braking.

For large displacement, when the drive has time to run up to its fixed maximum speed, the full displacement path must consist of three terms:

$$\Delta=S_r+S_b+S_{ms}$$

where S_{ms} is the path traversed at the constant maximum speed.

By using eqn (1) it is found that

$$\Delta=S_r(1+k)+S_{ms} \tag{3}$$

This expression shows that the device for determining the instant to start braking should be designed on the following principle: the path traversed during the run-up is continuously multiplied by $1+k$; to the value obtained at the instant of reaching the maximum speed the path traversed at that speed should continue to be added; and when the resultant quantity becomes equal to the initial error, then the command should be generated to start braking. It can readily be seen that expression (2) is a particular case of expression (3).

It has been assumed in the above discussion that the drive accelerations on run-up braking are constant, therefore their ratio k is constant also. But in practice k may vary between certain limits, which are not actually very wide; hence its maximum possible value is set into the computing device in question. With k smaller than the maximum, the last few millimetres of the path will be executed at a low speed, as has already been pointed out.

But in those cases where it is particularly vital to minimize the time of execution, self-adjustment may be introduced into the control system for the quantity k set into it. It is simplest to operate the self-adjustment according to the results of the completed execution, and for the self-adjustment criterion one should take the minimum both of the path length executed at low (creep) speed S_{cs} and also of the overrun path S_{r0} beyond the required point.

The ratio of the path $\Delta - S_{ms}$ to the run-up path is denoted by γ, and suffixes are given to all symbols as follows: 1 to indicate the previous action and 2 to indicate the next action. Then one can write

$$\Delta_1-S_{ms1}=\gamma_1\cdot S_{r1}$$

Since the creep speed is small enough one has:

$$\Delta_1=S_{r1}(1+k)+S_{cs1}+S_{ms1}$$

Since the aim of the self-adjustment is to establish the equation

$$\gamma_2=1+k$$

one gets

$$\Delta_1-S_{ms1}=S_{r1}\cdot\gamma_2+S_{cs1}$$

whence

$$\gamma_2=\frac{\Delta_1-S_{ms1}-S_{cs1}}{S_{r1}}=\frac{\Delta_1-S_{ms1}-S_{cs1}}{\frac{\Delta_1-S_{ms1}}{\gamma_1}}$$

Finally one has:

$$\gamma_2=\gamma_1\left(1-\frac{S_{cs1}}{\Delta_1-S_{ms1}}\right) \tag{4}$$

Employing an analogous argument for the case of overrun beyond the required point, one can write to a sufficient accuracy

$$\gamma_2=\gamma_1\left(1-\frac{S_{r01}}{\Delta-S_{ms1}}\right) \tag{5}$$

where S_{r01} is the overrun on the previous action.

Expressions (4) and (5) indicate the design logic for devices to give self-adjustment of the quantity k set into the system.

In control systems for high-power drives the presence of large inertia means that the current, and hence the motor torque, does not vary in a stepwise manner as shown in *Figure 1*, but much more slowly. This is evident from the oscillogram given in *Figure 2*, recorded for the motor of a blooming-mill screw-down mechanism.

For this reason, and also because considerable static loading is present, the motor speed, while varying with time in a roughly triangular law, lags behind the voltage during the run-up, and after the start of braking there no instantaneous reduction in speed; in fact it even goes on increasing for a certain time. Hence the ratio of the complete path to the run-up path required to the condition for optimal operation, which is a constant in the case of relatively low-powered drives with fixed characteristics, proves here to depend on the magnitude of the full action path itself, this dependence being of a complex non-linear nature. The relation $\gamma=f(\Delta)$ has been derived analytically for the screw-down mechanism of particular blooming mill and has then been checked on the mill itself, as shown in *Figure 3*. It should be noted that the curves for upwards and downwards motion are somewhat different, and the graph in *Figure 3* has been drawn from certain averaged-out values.

In the programmed-control system developed for the blooming-mill screw-down mechanism, the complete range of possible displacement values has been split into eight groups:

(1) less than 16 mm
(2) 16–32 mm
(3) 32–48 mm
(4) 48–64 mm
(5) 64–96 mm
(6) 96–128 mm
(7) 128–192 mm.
(8) greater than 192 mm.

The use of narrower intervals for small Δ is explained by the nature of the curve $\gamma = f(\Delta)$, whose slope gradually diminishes. The choice of the limits for the ranges was determined by the ease with which the given division could be engineered.

A special device forming part of the controlling part of the system automatically estimates the value of the initial error before each action, determines the group into which it falls, and sets up the mean value of γ corresponding to that group. The execution process itself proceeds similarly to that for the control of relatively low-powered motors, the nature of it being optimal in this case also by virtue of the fact that the run-up and braking accelerations are still constant and correspond to the maximum permissible torque value. It is only in the first two groups, for rarely met small displacements, that the excessively wide limits of variation of γ make it practically impossible to combine the optimum principle with accuracy requirements. Hence for the first group an action is used that is from start to finish at a lower speed equal to 10–12 per cent of maximum, while a limited speed is used for the second group.

If it is necessary to introduce self-adjustment of the quantity γ set into the control system, in this case it is evidently most desirable to apply the principle of altering the γ for a given group by the same increment at each repetition of a Δ corresponding to that group. A very complex installation would have to be designed in order to be able to apply the principle of self-adjustment of γ after the very first action.

Operation Algorithm of the Programmed-control System for the Screw-down Mechanism of a Blooming Mill

A system designed according to the above principle for controlling high-powered drives has two memory devices for rolling programmes:

(1) A static programme store (*SPS*) for long-term storage of fixed programmes specified according to the technological set-up for rolling at the works—40 programmes in all, with a maximum number of passes up to 23.

(2) A variable programme unit (*VPU*) for programmes that change often and are not stored in the *SPS*. There are two means for recording programmes on the *VPU*: (*a*) Manual recording using a telephone dial, and (*b*) Automatic recording of a rolling programme carried out under manual control by an operator. This allows one to use the system for automatically rolling a series of roughly identical unconditioned ingots for which no fixed programme is yet in existence. The operator uses his experience to roll the first of this series of ingots, the gap sizes set on the rolls being automatically recorded on the *VPU* during the rolling; the remaining ingots of the series are then rolled according to this recording.

As well as these methods of use, the *VPU* can also be connected to a computer calculating optimum rolling programmes. A single programme containing up to 35 passes may be recorded on the *VPU*.

In the developed system the size of the required gap between the rolls is given in the form of a ten-digit binary number, expressed in millimetres and equal to the distance from the initial point of a given position on the upper roll.

The operational algorithms for the systems of control from the *SPS* and from the *VPU* are basically identical; they contain the following operations or elements:

(1) Choice of operating régime (automatic operation from *SPS* or from *VPU*).

(2) Choice of the necessary programme (when working from *SPS*).

(3) Setting up the computing equipment to the initial position.

(4) Feeding in, from the programme store, of information on the given position for the upper roll.

(5) Determination of the actual position of the upper roll (interrogative operation) and computation of the initial error signal.

(6) Determination of the determination of rotation of the motor.

(7) Setting up the value of the coefficient γ.

(8) Start of operation.

(9) Attainment of maximum speed by the drive.

(10) Determination of the instant for braking to start, generation and execution of the relevant command.

(11) Transition of the drive to creep speed.

(12) Determination of the instant for stopping the drive, generation and execution of the relevant command.

(13) Transition from the given pass to the next one, all the operations from (3) to (13) then being repeated.

All the operations are carried out automatically except for (1) and (2) where the operator has to press the relevant push-buttons.

The automatic recording of a programme on the *VPU* with manual control follows this algorithm:

(1) Choice by the operator of the relevant régime.

(2) Setting of the upper roll to the required position.

(3) Setting up the computing equipment to the initial position.

(4) Interrogation of the measuring equipment to give the position of the upper roll, and translation of the resulting information into binary code.

(5) Transmission of the information to the *VPU*.

(6) On proceeding to the next pass, all the listed operations from (2) to (5) are repeated.

Operations (3), (4) and (5) are carried out automatically one after the other.

A programme can be set manually into the *VPU* using the telephone dial while the system is in operation from the *SPS*.

Block Diagram of Programmed-control System

The block diagram of the control system is shown in *Figure 4*. One of the fundamental elements of the system is a measuring unit *MU* of original design. It fulfils two functions: (1) on receiving an interrogation command it makes a single determination of the actual position of the upper roll, and gives out

a number of pulses equal to the gap between the rolls in millimetres, and (2) it signals the path traversed, giving out during the execution process a pulse for every millimetre traversed. In order to carry out these tasks the *MU* has two independent channels, one each for interrogation and for execution. It is linked to the screw-down mechanism by a synchro transmission. The interrogation operation takes place when the rolls are stationary and during the rolling of the metal.

A reversible binary counter *RC* is used to determine the magnitude of the initial error signal Δ, to derive the stop command and to record the rolling programme on the *VPU*. For convenience in the design of the computer section, the counter determines not Δ but its complement $\overline{\Delta} = C - \Delta$. Here $C = 1{\cdot}027$ is the counter capacity. The demand (D) in the form of a ten-digit binary number in direct code is introduced into the *RC* by a parallel means. Then the interrogate command is sent out, and the *RC* receives from the *MU* a number of pulses (Φ) corresponding to the actual gap between the rolls expressed in the complementary code $\overline{\Phi} = C - \Phi$. Hence the resultant number in the counter is $D + \overline{\Phi}$. Two cases arise:

(1) $D < \Phi$. In this case the upper roll must be displaced downwards by an amount $\Delta = \Phi - D$. In the counter one gets:

$$D + \overline{\Phi} = D + (C - \Phi) = C - (\Phi - D) = C - \Delta = \overline{\Delta}$$

(2) $D > \Phi$. In this case the upper roll must be displaced upwards by an amount $\Delta = D - \Phi$. So that the quantity $\overline{\Delta}$ should be derived in the counter also in this event, interrogation pulses must be added to D only till the counter is full; from that instant the switch *SW* puts the counter into the subtraction mode, and the arrival after this of the number

$$(C - \Phi) - (C - D) = D - \Phi = \Delta$$

of pulses from the *MU* gives in the counter the quantity

$$C - \Delta = \overline{\Delta}$$

During the execution the counter always operates in the addition mode. When it receives from the *MU* a number of pulses equal to Δ, it overflows

$$\overline{\Delta} + \Delta = C - \Delta + \Delta = C$$

and gives a pulse from its last digit that is used in the command unit *CU*1 to generate the 'stop' command.

A straightforward logic designed into the command unit *CU*1 generates the command 'up' or 'down' according to whether the binary counter has overflowed or not during the interrogation process. These commands are passed to the logic unit for the drive control.

A transfer register connected to the reversible counter and repeating all its actions serves for the transfer of the quantity $\overline{\Delta}$ derived in the counter to the device for determining the coefficient γ and to the non-reversible binary counter *BC* that serves to determine the instant for giving the command to start braking. It is also used when a programme carried out by a rolling operator is being recorded on the *VPU*. In this event the reversible counter is put into the read-out mode, and then interrogation of the *MU* is carried out. As a result one obtains in the counter and the transfer register the magnitude of the gap between the rolls in direct code:

$$C - \overline{\Phi} = C - (C - \Phi) = \Phi$$

This information is read over in the transfer register and transferred to the *VPU* by a parallel means.

The frequency divider *FD* serves to generate the various values of the coefficient γ. It consists of a normal binary counter to the cells of various digits of which are connected the inputs of switches *K*4–*K*10. Thus, for example, if the outputs of the first, third and seventh digits are connected to any switch, then when 128 pulses arrive at the input of the frequency divider from the *MU*, $64 + 16 + 1 = 81$ pulses will reach the switch. If the output of this switch is connected to the input of the third digit of the braking binary counter, then evidently the coefficient $\gamma = 81/128 \cdot 4 \simeq 2{\cdot}53$.

The role of the device described later for determining the quantity γ consists in opening whichever of the switches *K*4–*K*10 will set up the required value of γ for a given Δ.

Since $\overline{\Delta}$ has already been recorded in the braking counter, therefore when Δ/γ pulses have been received from the *MU* the counter becomes full and its last digit gives out a pulse that is then used in the command unit *CU*1 for forming the braking command, realized by the drive control logic unit.

If, before the braking counter becomes full, the drive has time to run up to its fixed maximum speed, then from that instant all the switches *K*4–*K*10 are closed, and by opening switch *K*3 the number of pulses originated by the measuring unit is passed to the first digit of the counter. This carries out the logic for determining the instant to start braking, as already described.

The next section describes the devices for automatically limiting the maximum drive speed and for transition to creep speed during the braking process. The path length traversed at creep speed is 3–5 mm.

The system is started up automatically by a photoelectric relay system at the instant when the metal leaves the rolls. But because the motor has a delay in starting of 0·6 sec, a corresponding advance must be introduced. This is achieved by a special assembly that indirectly measures the speed of the metal and generates a pulse to start the system calculated so that the drive starts at the instant when the metal leaves the rolls. This assembly is *not* shown in *Figure 4*.

The system also contains a number of elements that carry out various logical functions required for the sequencing of the operations, for their automation, etc. In particular, a photoelectric relay unit is mounted on the mill for automatic drive starting.

The system provides for control of the most responsible operation—the stopping of the drive at the correct time. For this purpose the reversible counter is duplicated. The outputs of both counters are fed to a special control logic unit. If overflow pulses are not generated simultaneously by both counters, this unit gives out both a stop pulse and a fault signalling pulse; if both overflow pulses arrive at once, it generates only a stop pulse.

Certain Basic Elements of the Control System

(1) *Electric Motor Drive and its Control*

The electric motor drive for the blooming-mill screws is designed as a generator-motor system. A 375 kW d.c. generator

powers the two 180 kW screw-down motors connected in series, and is controlled by a 4·5 kW amplidyne.

The drive must provide for the execution of a prescribed path according to the optimal speed curves given in *Figure 1*. In this connection the following requirements are placed on the drive:

(1) In order to obtain the maximum response speed, the motor current must be held equal to the maximum permissible during run-up and braking.

(2) Limitation of the maximum rotation speed of the motor is necessary.

(3) During the braking process an automatic transition must be ensured to the creep speed $n = n_{min}$.

(4) Heavy braking is necessary when the drive is finally stopped from creep speed.

The layout of a drive satisfying these requirements is shown in *Figure 5*. The control winding $W1$ of the amplidyne is connected tothe output of a three-state semiconductor trigger circuit which receives control pulses from the drive control logic unit. The run-up and braking of the drive take place at an invariable value of motor current $I_m = (I_m)_{max}$, which is achieved by the use of strong negative current feedback in the armature circuit (feedback winding $W4$), with a feedback gain of 8–10. For large error signals, when the voltage at the generator terminals reaches its maximum value, depending on its polarity one of the stabilovolts *ST* strikes. This causes the maximum-speed relay *RMS* to operate and apply the generator voltage to winding $W3$. The current flowing in this winding sets up a negative feedback that limits the generator voltage and consequently the motor rotation speed.

The creep speed is obtained by means of the twin-winding relay *RCS*. This relay is operated at the start of the execution by one of the windings being energized. At the start of braking this winding is de-energized, and the relay is held on only by the action of the second winding, which is energized from the generator output voltage; as this voltage falls in consequence of the braking process, the relay drops out and causes a strong negative feedback to be applied, which together with the change in the polarity of the current in the amplidyne control winding sets up a speed that is about 10 per cent of the maximum. Efficient braking from this speed on stopping is achieved by the self-damping of the generator on the removal of the control action from the control winding $W1$.

As stated above, the drive control equipment consists of a three-state power trigger circuit whose output is connected through a balanced semiconductor amplifier to the amplidyne control winding $W1$.

In order to obtain the required variation in the control action, pulses must be supplied to the appropriate inputs of the trigger circuit. The order of application of the pulses depends on the direction in which the upper roll has to be displaced; it is developed by the drive control logic unit.

Signals are fed by six channels to the input of this unit from the digital control system and the drive system. These commands are as follows: selected direction of motion (up or down), clearance to start, braking, transition to creep speed, and stop. From these commands the logic circuit derives the signals that go to the appropriate trigger circuit inputs.

(2) *Static Programme Store*

The static programme store *SPS* (*Figure 6*) is a matrix memory device in which binary numbers forming a programme are recorded by means of networks of semiconductor diodes. It consists of a distributor, a programme unit and a numerical unit.

The distributor (see the bottom line of *Figure 6*) sequentially sends out a read pulse to the programme unit (second line of *Figure 6*) in accordance with the sequence of passes making up each programme; it is a device without moving parts that switches from pass to pass. The maximum number of passes in the programmes is 23, and so the distributor has 23 digits (23 ferrite-transistor cells).

The programme unit consists of 23 ferrocart programme cores, each of which has one primary winding connected to the distributor and 40 secondary windings (one for each fixed programme). The secondary windings of all the cores for a given programme are all connected at one end to a common bus, while the other ends go to the diode numerical matrices. Selection of the required programme is made by connecting one or other of these secondary-winding bus-bars to the output bus (+ on the diagram). Thus the operator needs only to press a button on the control desk to select the required programme.

(3) *Variable Programme Unit*

A fundamental element of the *VPU* is its store *ST*, consisting of a ferrite matrix on which 35 ten-digit numbers can be recorded. Each core of the matrix has four windings: erase (reset), carry-in of numbers, write (also serving as read-out winding), and output (*Figure 7*). The carry-in and output windings of the ferrites for the same digit are connected in series (35 ferrites each); the write and read-out windings of all the ferrites for a given number (pass) are also connected in series (10 ferrites each).

The operation of the store is based on the well-known Cambridge principle. But the design logic and circuit are original and very simple.

For the recording of a number in the store, pulses are applied to the input shaping circuits for the appropriate digits. At the same time an activation pulse is applied to the distributor, 35 of whose cells have their outputs connected to the corresponding write and read-out windings. *Figure 8* shows the form of the pulses generated by the distributor and shaping circuits, and also their relative timing. The sense of the current corresponding to the top part of the pulses is for read-out. Hence, as is clearly seen from *Figure 8*, the superposition of the two magnetizations on the ferrite at the start only confirms the absence of recording, while later on (when the bottom parts of the pulses in *Figure 8* coincide) a 1 is written.

When reading out numbers, an activation pulse is supplied each time to the distributor, and the pulse coming from it performs the read-out. So as to regenerate the read-out number, feedback is taken from the output shaping circuit of each digit to the input shaping circuit for the same digit, resulting in the appearance of a pulse from the input shaper almost at the same instant as a register pulse appears; but the relation of the initial parts of these pulses is such that this attenuates the read-out pulse only negligibly. A coincidence of the magnetizations (the lower halves in *Figure 8*) brings about regeneration of the number—its re-recording. By this means the recorded pro-

gramme may be reproduced a practically unlimited number of times.

As already pointed out, the recording of a rolling programme carried out by an operator under manual control is achieved by means of the reversible counter included in the system. There is a special original device for the manual recording of programmes. A number is dialled on a somewhat modified telephone dial, taking its digits in sequence one after the other. To record the number 253, for example, 2, 5 and 3 are dialled in sequence, while to record 72 one dials 0, 7 and 2 in sequence, etc. The dial has two contact systems: one for numerical pulses and one for control pulses, which are fed out on separate channels. The dial is designed so that when one dials zero only two control pulses are generated (one each for clockwise and anticlockwise rotation of the dial); when one dials 1 there is one control pulse, one number pulse and then another control pulse, etc. The control pulses thus generated serve to activate the six-digit distributor controlling the recording system. The outputs of its cells control switches in such a way that the first switch (hundreds) is open at the instant when the number pulses come through for the first digit of the number to be recorded, the second switch (tens) for the second-digit pulses, etc. These pulses are passed from the switches to a binary counter that serves to form the binary code for the number (*Figure 9*).

This device works on the principle of introducing pulses into the digits of the binary counter in such a way that the sum of their values equals the number of pulses received. For example, since the number 100 has the form 1100100 in binary code, for every pulse arriving from the first switch (hundreds) one pulse is put into the third, sixth and seventh digits of the binary counter; so as to avoid disruption of the computation in the event of digits being carried from lower to higher columns, these pulses are supplied not simultaneously to all three of the digits mentioned, but spaced by a time delay which is enough to allow the carry to take place.

After the dialling of the third figure is complete, the final control pulse causes a pulse to be sent out from the output of the sixth cell of the distributor, which in its turn brings about the transfer into the store of the number formed in the counter, followed by the preparation once more of the first cell of the distributor. This makes it possible to dial numbers continuously one after the other. The correctness of the dialling may be checked on a visual indicator of dialled numbers, which uses three dekatrons. The operator has the facility of erasing a number when necessary by pressing a button (shifting the distributor backwards by one cell), and of then recording it again.

(4) *Coefficient Selection Unit*

Figure 10 shows the block diagram of this unit. As has already been stated, the quantity γ is chosen in accordance with the value of Δ, while the whole range of variation of Δ is divided into eight groups.

The unit contains three basic elements:

(1) Ferrite assembly (top line in *Figure 10*). These ferrite cores serve for the estimation of the value of Δ, and are connected into the lines for transferring $\overline{\Delta}$ from the reversible counter to the braking counter. There are eight of them altogether, and on them are written the eight highest digits of $\overline{\Delta}$.

(2) Switch assembly (middle line in *Figure 10*). The switches serve to control the lines for various values of γ. Since, as was pointed out, the execution for one group of Δ (from 0 to 16mm) is carried out from start to finish at creep speed, the number of switches is one less than the number of groups, i.e., seven.

(3) Transformer assembly (bottom line in *Figure 10*). The transformers have ferrocart cores, and each serves for the setting up of a certain value of γ. For this purpose each core has several primary windings, to which are connected the outputs of those digits of the frequency-divider that are required to give the necessary value of γ. The secondary (output) winding of the core is connected to the input of the corresonding switch.

The estimation of the value of Δ is based on the following principle:

Since what is written on the ferrite cores is not the value of Δ itself but its complement $\overline{\Delta}$ w.r.t. 1024, the following picture is obtained for various groups of values of $\overline{\Delta}$:

(1) $\Delta \leq 16$: 1's are written in all the digits from the fifth upwards; one or more of the cores for the first four digits contains a 0.

(2) $16 < \Delta \leq 32$: 1's are written in all the digits from the sixth upwards; the core for the fifth digit contains a 0.

(3) $32 < \Delta \leq 64$: 1's are written in all the digits from the seventh upwards: the core for the sixth digit contains a 0.

(4) $64 < \Delta \leq 128$: 1's are written in all the digits from the eighth upwards: the core for the seventh digit contains a 0.

(5) $128 < \Delta$: one of the digits from the eighth upwards contains a 0.

Making use of the above, the device is designed in the following manner.

Immediately after $\overline{\Delta}$ has been recorded on the ferrite cores, it is read out with polarity such that those cores containing 0's give pulses in their output windings. After amplification by triodes, these pulses are passed to windings for opening switches corresponding to these cores. So that several switches should not open all at once, the opening winding for each is connected in series with the shut-off windings for all the switches corresponding to cores of lower digits. Thus each time only one switch opens, corresponding to the core of the highest digit in which no 1 is written.

To consider the means by which certain of the above intervals are split into two, the interval $32 < \Delta \leq 64$ is taken as an example. This is split into the two parts (1) $32 < \Delta \leq 48$ and (2) $48 < \Delta \leq 64$.

In addition to the conditions for this interval, an extra one will exist for the first half—the presence of a 1 in the fifth digit: while for the second half it will be the absence of a 1 in the fifth digit. In order to control the satisfaction of these conditions, an extra ferrite core is connected in the transfer line for the fifth digit, and on read-out it gives a pulse in its output winding when a 1 is present on it. This pulse is amplified by a triode and closes a switch corresponding to the band $48 < \Delta \leq 64$, and in spite of the fact that on all occasions when $32 < \Delta \leq 64$ the switches for both parts of this interval receive opening pulses, nevertheless for $32 < \Delta \leq 48$ only the switch for this band is open. For $48 < \Delta \leq 64$ the main ferrite core for the fifth digit closes this switch, and only the switch for the band $48 < \Delta \leq 64$ remains open.

The other intervals are split up in a similar manner.

(5) *Start Pulse Advance System*

As was observed earlier, the task of this system is to generate a pulse for starting the drive at a certain roughly constant time (about 0·5–0·6 sec) before the metal comes out of the rolls. This calls for an estimate of the speed of motion of the ingot up to that instant.

For this purpose two photoelectric relays are mounted on either side of the rolls, at distances of 0·5 and 1 m from the plane of the axes of the rolls. These relays control the mode of operation of a special reversible counter and a fixed-frequency generator supplying pulses to this counter at frequency f.

Let t be the time of advance, n_1 the number of pulses which reach the counter from the instant of obscuration of the first photocell until the obscuration of the second, n the number of pulses that should reach the counter from the instant of obscuration of the second photocell until the generation of the start pulse, and C the counter capacity.

Then, if for practical purposes the assumption that at the end of its passage the velocity of the ingot is constant may be taken as acceptable, one must have:

$$n = n_1 - t \cdot f$$

The constant quantity $t \cdot f$ is first set into the counter, which is put into the read-out mode. At the instant when the first photocell is obscured, a pulse switch is opened, and until the instant of obscuration of the second photocell n_1 pulses enter the counter. The following quantity is got in the counter:

$$C + t \cdot f - n_1$$

From this instant the counter is switched into the addition mode, and when $n = n_1 - t \cdot f$ pulses have entered it a pulse appears from its last digit, which is in fact used for starting the drive.

(6) *Measuring Unit*

As has been stated, this unit has two channels: interrogate and execute. Position transmitters with oscillatory circuits are used for both channels, and both are equipped with discs having tooth-like perforations round their edges. The disc for the execution channel is linked by a synchro transmission to the screw-down mechanism, while the disc of the interrogate channel is continually rotated by a small motor. Pulses appear in the channels when the teeth of the discs enter the inductors of the corresponding sensor circuits. The instant for starting the count of interrogation pulses is determined by a transmitter of the same type, for which there is one special tooth on the periphery of the disc. The instant for stopping the interrogation is determined by a special electromagnetic transmitter, which gives out a pulse when a magnetic circuit linked by its parts to both discs is completed. The measuring circuit is designed entirely from elements without contacts.

Conclusion

In conclusion it should be noted that the tests of the programmed-control system for the screw-down mechanism have given positive results: the error in setting the upper roll did not exceed 1–2 mm, while the time of operation of the screw-down mechanism over the complete programme was less than the time of operation under manual control.

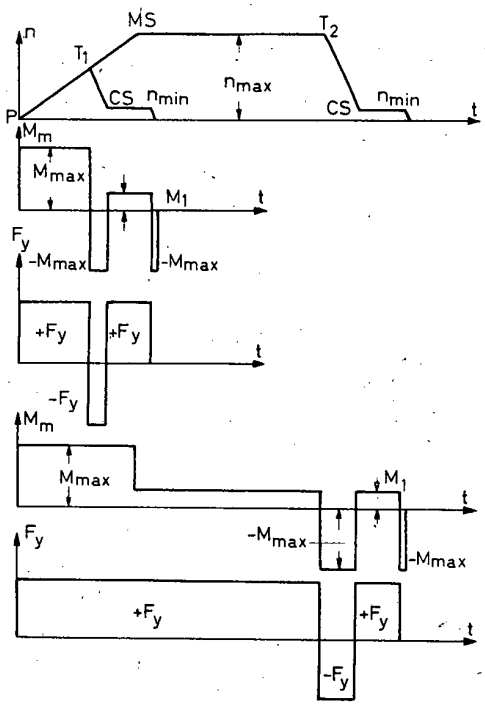


Figure 1

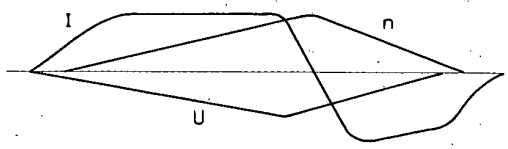


Figure 2

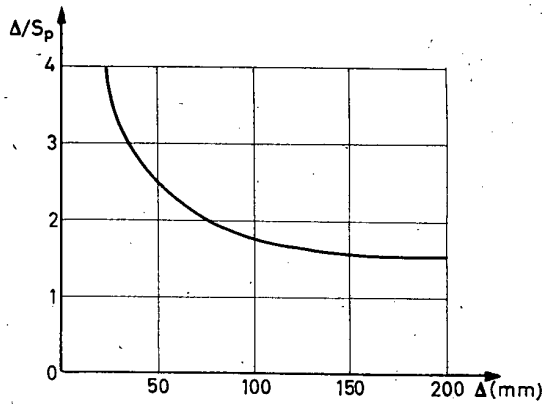


Figure 3

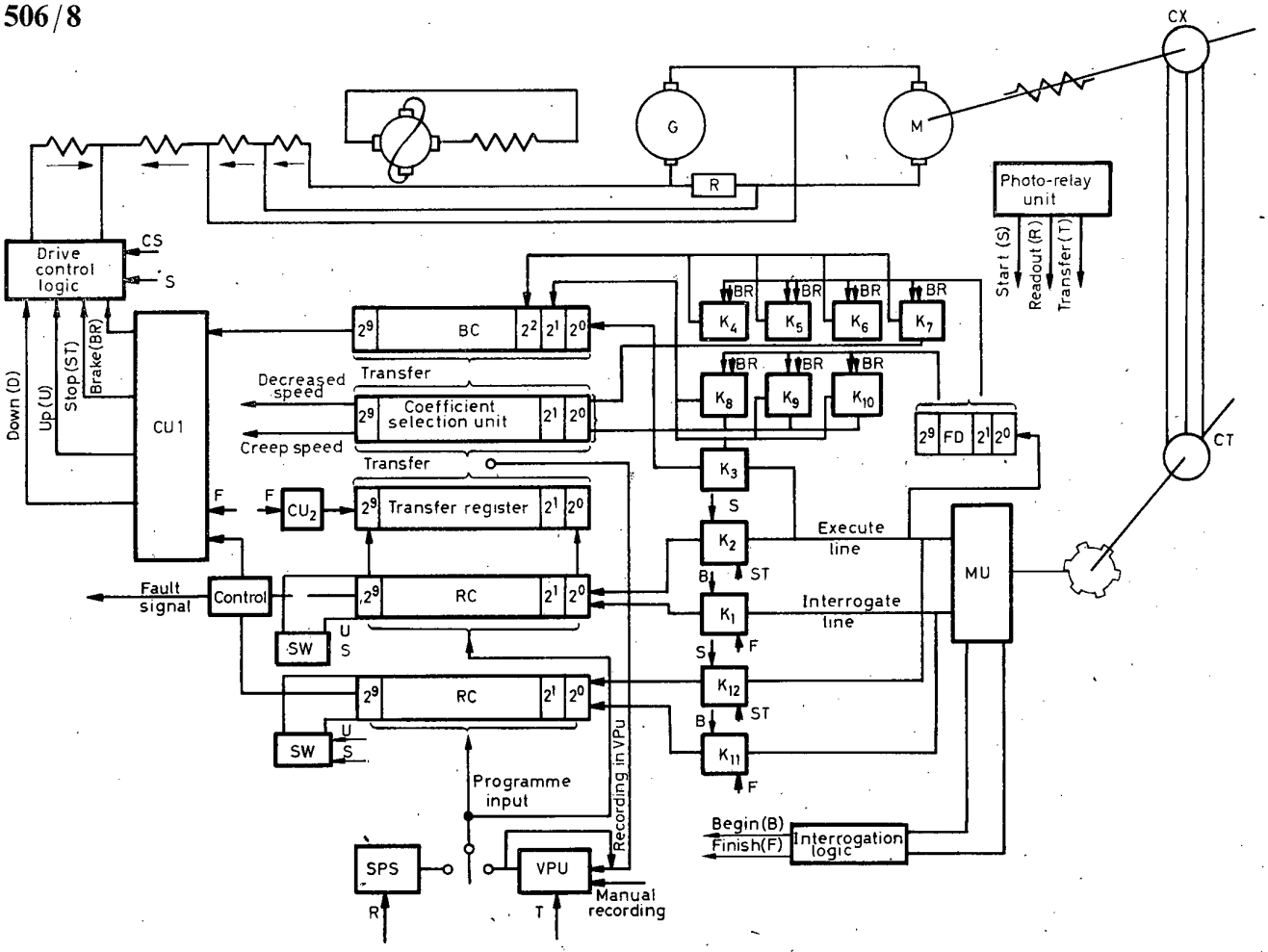


Figure 4

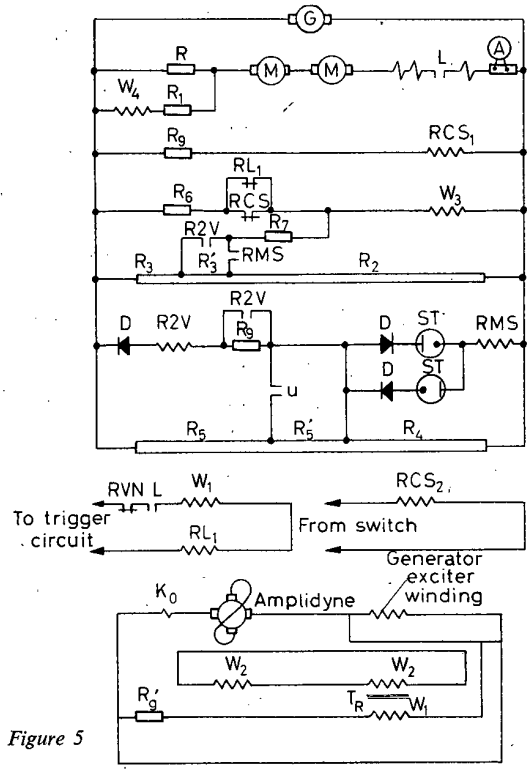


Figure 5

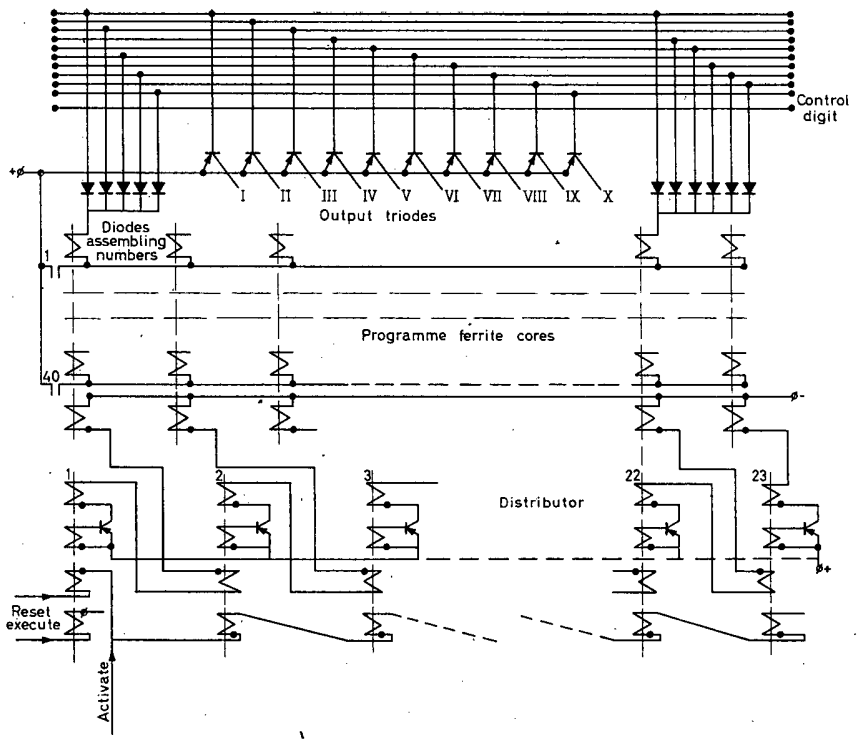


Figure 6

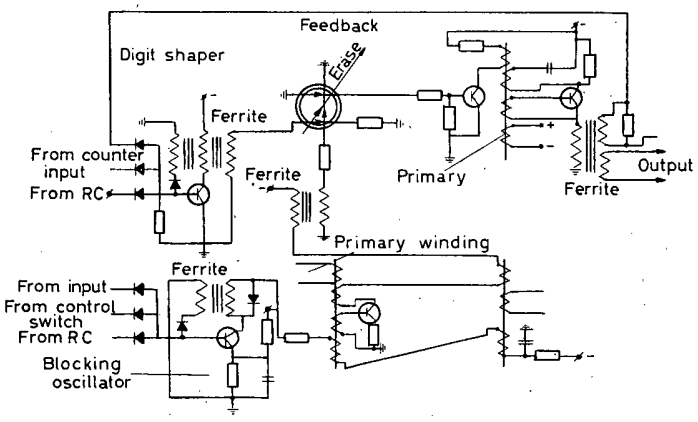


Figure 7

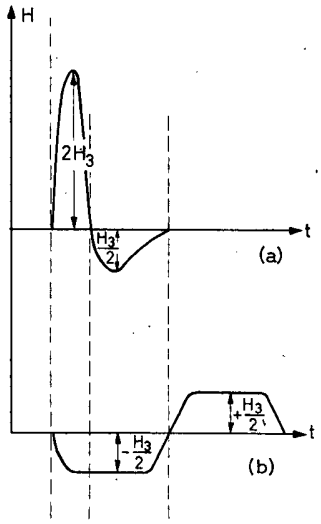


Figure 8

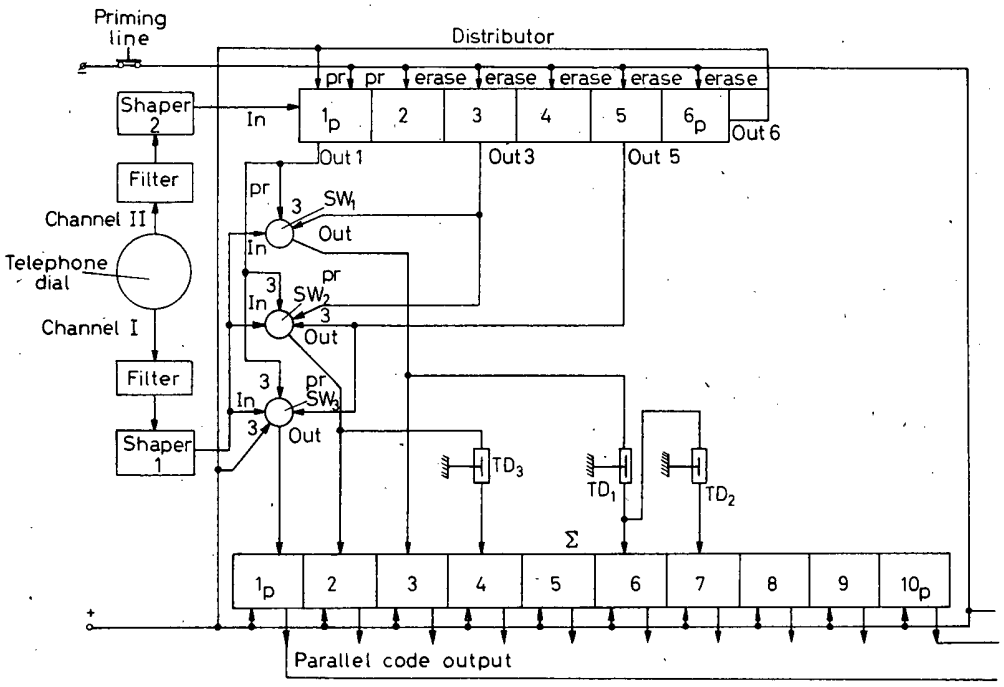


Figure 9

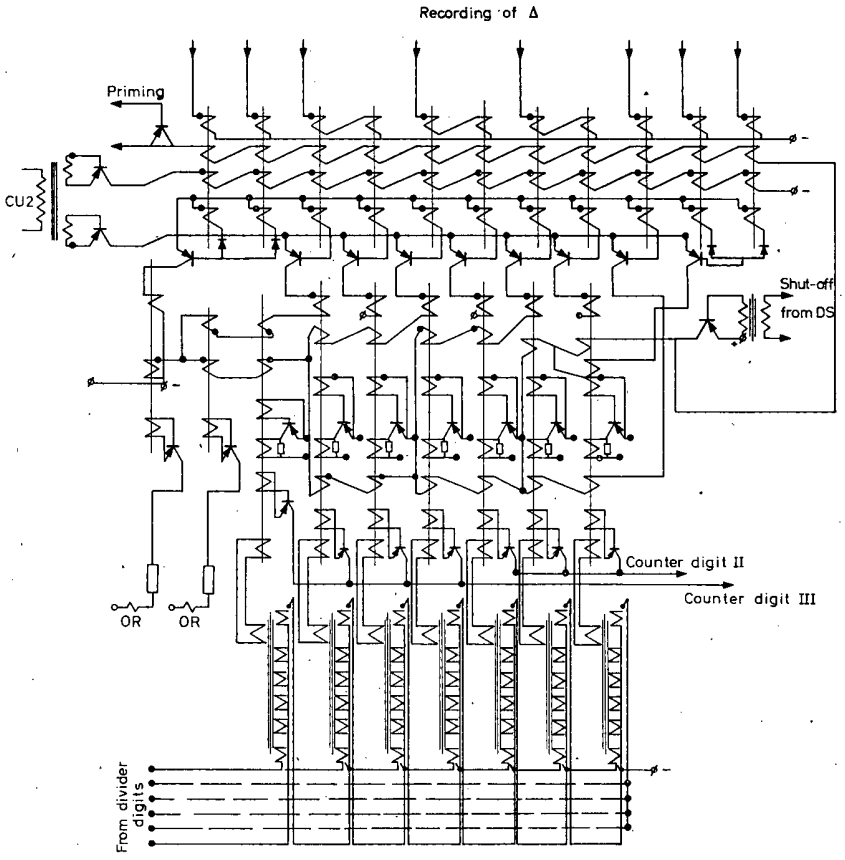


Figure 10

Study of Industrial Production of Polyethylene under High Pressures, and of the Automatic Control of the Process

B. V. VOLTER

The process of ethylene polymerization under high pressures represents one of the chemical engineering processes which is most difficult to control. Its characteristic features are high pressure, frequent explosions, considerable variations in the output of the reactor and quality of the product. The ordinary systems used for automatic stabilization of parameters do not guarantee normal progress of the reaction for this process. Therefore, a satisfactory solution of the problem of automatic control of the process for industrial manufacture of polyethylene can be solved only by the synthesis of a special system of control on the basis of data of experimental and theoretical study of the process.

The investigation of the process and the study of the automatic control systems were conducted on an experimental industrial reactor, which consists of a tube (see *Figure 4*) 50 m long having an internal diameter of 16 mm, with an outer water jacket for the initial heating of the reactive mixture and for the removal of heat in the zone of reaction. Gaseous ethylene at a pressure of 1,500 atm is continuously supplied to the reactor. The process of polymerization takes place at a temperature of about 200° C.

The Static Characteristics of the Process

For the study of the static behaviour of the process the method of non-linear multiple correlation was used[1]. The relationship between the output of the reactor and the basic parameters was represented in the form of a product of functions of single parameters.

$$Q = f_1(P) \cdot f_2(t) \cdot f_3(O_2) \cdot f_4(V)$$

where $f_1(P)$ is the function of pressure, $f_2(t)$ the function of temperature, $f_3(O_2)$ the function of oxygen concentration, and $f_4(V)$ the function of gas supply to the reactor.

Each one of these functions was represented in the form of a polynomial

$$f_i(x_i) = a_i + bx_i + cx_i^2$$

The results of periodic measurements of output and of other parameters of the process were used as the initial data for the calculation. By the construction of the correlational fields and by developing the regression curves for each parameter, the coefficients of all functions $f_i(x_i)$ were determined. The general formula for the output of the reactor has the form

$$Q = 13 \times 10^{-8}(-211 + 0{\cdot}33\,P - 1{\cdot}16 \times 10^{-4}\,P^2) \cdot (t - 112)(O_2 - 55)(V + 587)\ \text{kg/h}$$

Its verification on an industrial installation gave quite satisfactory results.

The Stability of the Polyethylene Polymerization Reaction

The difficulty of controlling the process is aggravated by the risk of a reaction taking place which would result in the decomposition of ethylene into carbon, hydrogen and methane, which develops very rapidly and is accompanied by the liberation of large quantities of heat. When the signs of the risk of decomposition appear it is necessary, almost instantly, to reduce the pressure in the reactor or to diseharge the contents of the reactod into the atmosphere. If the decomposition of ethylene cannot be prevented then, instead of the expected valuable product, soot is obtained. Each decomposition is followed by a prolonged stoppage of production, which is needed to test the pressure tightness of the equipment, for the removal of soot from the inner surfaces of the reactor and for the carrying out of other usual operations. All this causes great production losses.

The study of the causes of the ethylene decomposition reaction and the development, on this basis, of methods and means for its prevention represents an essential problem. By using a special equipment it was possible to record several interesting moments in the operation of the reactor, which provide a possible explanation for one of the basic causes of the decomposition. Recordings showed that very often the normal progress of the process is disrupted by a sudden increase in pressure, reduction in gas consumption and by an abrupt increase in temperature. Such a sudden disruption of the operating conditions may be explained by the formation of polyethylene blockages in the reactor tube.

Rapid reduction of pressure in that case leads to the elimination of these blockages and to the slowing down in the reaction development. If the pressure is not reduced in good time decomposition reaction unavoidably develops.

It was also possible to record the picture of the explosion (*Figure 1*). In the example given the operator was unable to prevent the explosion by the reduction of pressure and, therefore, the contents of the reactor were discharged into the atmosphere. The decomposition of ethylene occurred in the reactor, as was indicated by a black cloud of soot discharged from the reactor. From che diagram it is also evident that there was an increase in pressure and temperature which was accompanied by an abrupt reduction in gas consumption.

For the elimination of the polyethylene blockages it was proposed that at the very beginning of their formation forced oscillations in pressure should be induced. An oscillator, specially developed for this purpose, fully justified itself in operation. Another means of preventing these blockages is to incerase the gas supply to the reactor.

The measures indicated did not lead to a complete elimination of decompositions, although they became less frequent

but all the same they took place. This circumstance points to the instability of the polymerization reaction itself.

The first experiment for the investigation of the stability of the polymerization reaction of ethylene under high pressure was undertaken by Hoftyzer and Zwitering[2]. Having constructed the material and thermal balance equations for an elementary part of the reactor, the authors obtained two non-linear differential equations in a dimensionless form:

$$\frac{dy}{dz} = y e^{-\frac{(1+u)}{x}} V(y_0 - y) \quad (1)$$

$$\frac{dx}{dz} = \frac{1}{2} y^{\frac{1}{2}} e^{--} + V(x_0 - x) - VW(x - x_w) \quad (2)$$

where y_0, y are the inlet and outlet concentrations of the initiator, x_0, x the inlet and outlet temperatures, x_w the reactor wall temperature, z the time, u the parameter, which determines the activation energy, and v, w the constant coffiecients.

Using Liapunov's method[3], the stability of the state of equilibrium was investigated by the linear equations of first approximation:

$$\frac{dX}{dz} = a_{11}X + a_{12}Y \quad (3)$$

$$\frac{dY}{dz} = a_{21}X + a_{22}Y_i \quad (4)$$

According to Liapunov's method, the stability of the equilibrium state x_s, y_s of a non-linear system of the second order is determined by the following Routh–Hurwitz conditions:

$$d_f = a_{11}a_{22} - a_{12}a_{21} > 0 \quad (5)$$

$$d_a = -a_{11} - a_{22} > 0 \quad (6)$$

By equating the left sides of these inequalities to zero, the authors determined the boundaries of the region of the stable equilibrium states for an area of parameters x_0, y_0 for the different values of x_0. They arrived at two interesting results: (1) the system can have five states of equilibrium; and (2) the industrial reactors are operated in a region where condition (5) is satisfied, but where condition (6) is not satisfied.

The investigation of eqns (1) and (2) terminates at this point, and Routh and Hurwitz proceed to the study of the system of control. However, in the author's opinion the study of the reactor itself was left unfinished.

First of all, the question arises: is the region of unstable states of equilibrium the region of decompositions? The instability of the state of equilibrium may lead either to a rapid increase in temperature or to stable temperature oscillations. From the theory of oscillations[3] it is known that in non-linear systems self-oscillations are possible—stable periodic oscillations which in the absence of external disturbances are periodic in character. The phase picture of self-oscillating systems contains at least one isolated closed trajectory—the limiting cycle. If the limiting cycle is stable, then the state of equilibrium embraced by this cycle will be unstable. It follows from this that the problem of stability of the polymerization reaction of ethylene is closely associated with the problem of self-oscillations. However, before undertaking any theoretical investigation of self-oscillations, it is necessary to be convinced about the practical expediency of this. In other words it is necessary to possess the experimental material which would confirm the possibility of self-oscillations in an actual process of ethylene polymerization.

Self-oscillations of the Ethylene Polymerization Reaction

The possibility of the occurrence of periodic oscillations in chemical systems has been known for a long time. Andronov[4] indicated that under certain conditions in chemical systems, just as in other (mechanical, electrical, etc.) systems, continuous oscillations, inexplicable in principle by the linear theory may occur. Recently, a large number of works devoted to the experimental and theoretical study of the periodic chemical reactions, were published. A detailed outline of these investigations is given in the work of Salnikov[5].

The observations made on the process of ethylene poly merization in a tube reactor shows that the process takes place under the conditions of abrupt oscillations in temperature, reactor output and quality of the product.

In the manual control of the process it was possible to explain these oscillations by the instability of pressure and oxygen content in the mixture, by the change in the gas supply and by other causes, i.e., it was possible to consider that these changes in the process represent imposed changes. For us it was quite unexpected to find that the automatic stabilization of basic disturbances had very little effect on the progress of the process. The oscillations in temperature, output and quality, as before, remained considerable. This very fact suggested that the process has its own inherent internal rhythm, determinable only by the properties of the system, and not by the external disturbances, i.e., that self-oscillations are characteristic of the process.

In *Figure 2* are given the diagrams of recordings of pressure and temperature along the length of the reactor, from which it is seen that the temperature oscillates constantly, under which conditions the period and the amplitude of these oscillations change along the length of the reactor. (The term 'amplitude' is used conditionally, since the oscillations are not harmonic.) The increase in the period of oscillations at the end of the reactor is clearly seen. At point No. 13 the period amounts approximately to 15 min, at the fourteenth point it is already 20–25 min, and at the last point it exceeds half an hour. The amplitude of temperature oscillations along the length of the reactor also increases continuously, and at the last point it reaches 30–40° C. The experiments were carried out in the presence of forced pressure oscillations having an amplitude of 70 atm and a period of 2·5 min. These oscillations are recorded on the pressure diagram. The period of these oscillations is 10 times less than that of the natural temperature oscillations. The pressure oscillations are reflected in the temperature, although not very appreciably. They are, for instance, superimposed on temperature oscillations and do not alter the general picture at all.

It is possible to note yet another peculiarity in the behaviour of the process: the temperature oscillations at the first points akong the gas flow are not reflected at all in the oscillations at subsequent points. This is as if each part of the reactor represented an isolated oscillating system having its own period and amplitude. This, at first glance, contradicts common sense, since the gas moves through the reactor at a high velocity, and it would be more natural to expect an interdependence in the behaviour of temperatues in neighbouring points.

It will be assumed now that the temperature oscillations are

conditioned by external disturbances. Then, however, a large number of inexplicable questions is raised. First of all, what force should these disturbances have if pressure oscillations of 70 atm are hardly reflected in the temperature? Why do these external forces cause temperature oscillations, having quite different periods, along the length of the reactor? Why are these disturbances more pronounced at the end of the reactor than at the beginning? Finally, why, in general, should these external disturbances cause almost periodic temperature oscillations? All these questions, in our opinion, are inexplicable when taking into account only the external disturbances. Therefore, the deduction that temperature oscillations are explained by the internal oscillational nature of the process, i.e. by the self-oscillations of the reaction, is more convincing.

The explanation of all peculiarities in temperature behaviour in the reactor requires a detailed study of the mechanism of the reaction. In this paper only hypothetical reasons concerning some questions are given.

The period of temperature oscillations may increase as a result of a decrease in the concentration of the initiator at the end of the reactor. An increase in the amplitude of oscillations is probably determined by an increase in the viscosity of the mixture as a result of polymer formation. It is known that an increase in the viscosity of the reactive mixture usually tends to inhibit the chain break-up reaction, but that it has no effect on the chain growth. Therefore, at the end of the reactor longer polymer chains should be formed, and since liberation of heat is determined by the chain-growth reaction, then this also leads to an increase in the amplitude of temperature oscillations at the end of the reactor.

The fact that the temperature oscillations at the neighbouring points are not correlated among themselves can be explained by the action of the reactor wall. It is generally known that the wall, in some reactions, plays a big role. Very often the break in the chain reaction occurs on the wall, and in other reactions the wall also participates in the initiation of the chain. Semenov[6], for example, points out that the molecule of oxygen on the reactor surface can enter into the reaction $V + O_2 \rightarrow VOO$, as a result of which a powerful peroxide radical VOO is formed on the surface. The latter reacts readily in the presence of hydrogen with the initial substance, for example RH, giving a surface peroxide compound VOOH and radical R. For such reactions, the liberation of heat not in space but on the wall, is characteristic; and if the reaction is irreversible, then in the course of the process the wall is covered by a chemisorptive layer and its initiating action ceases.

If a similar picture could be built up for the ethylene polymerization reaction (which of course requires a special proof), then it is possible to visualize that the mechanism of self-oscillations of the reaction will be as described. The progress of the reaction leads to an increase in temperature, but the coating of the wall by the chemisorptive layer of polyethylene molecules leads to the damping of the reaction and to a reduction in temperature. Gradually with the gas flow the polymer is washed off the walls, the reaction develops again, the temperature increases, and so on. Naturally, because of such a mechanism the reactor will consist of a large number of self-oscillating systems, distributed along the length of the reactor. Under these conditions the temperatures at the neighbouring points will not be mutually interconnected.

It is possible to put forward a number of other self-oscillating models of the process. In view of the strongly pronounced exothermic nature of the process it is most likely that the self-oscillations are thermo-kinetic in character, in which case the interaction between the heat removal system and the reaction leads to stable temperature oscillations. Similar oscillations were studied for the first time by Frank-Kamentskii[7].

Study of the Thermo-kinetic Model of Reaction

The rate-of-reaction equation for the polymerization of ethylene may be represented in the form:

$$-\frac{dM}{dt} = A e^{-\frac{E}{RT}} I^{\frac{1}{2}} M$$

where M is the concentration of monomer, I the concentration of initiator, E the activation energy, R the gas constant, T the temperature, A the pre-exponential multiple, and t the time.

On the basis of the rate-of-reaction equation it is possible to construct for an elementary section of the reactor the material and thermal balance equations.

$$\frac{dM}{dt} = -A e^{-\frac{E}{RT}} I^{\frac{1}{2}} M + \frac{G}{V}(M_0 - M) \tag{7}$$

$$C\rho \frac{dT}{dt} = VQA e^{-\frac{E}{RT}} I^{\frac{1}{2}} M - Sh(T - T_v) + G\rho C (T_u - T) \tag{8}$$

here G is the gas supply, V, S the volume and surface of the reactor section under consideration, M_0 the monomer concentration in the initial mixture, Q the thermal (calorific) effect of the reaction, C the specific heat of the mixture, h the heat-transfer coefficient, ρ the density of the mixture, and T_u, T_v the temperature of the mixture and temperature of the reactor walls. By denoting that

$$d = \frac{G}{V}, \alpha = \frac{Sh + G\rho C}{V}, T_0 = \frac{S_v T_u + G\rho C T_v}{Sh + G\rho C}$$

the system may be reduced to the following form

$$\frac{dM}{dt} = -A e^{-\frac{E}{RT}} I^{\frac{1}{2}} M + \alpha(M_0 - M) \tag{7a}$$

$$C\rho \frac{dT}{dt} = Q e^{-\frac{E}{RT}} I^{\frac{1}{2}} M - \alpha(T - T_0) \tag{8a}$$

If it is assumed that the concentration of the initiator is constant and if dimensionless variables $x = QR/C\rho E$, $y = R/(E)T$, $\tau = A I^{\frac{1}{2}} t$ are introduced, then the material and thermal balance equations will be

$$\frac{dx}{d\tau} = -x e^{-\frac{1}{y}} + \beta(x_0 - x) \tag{7b}$$

$$\frac{dy}{d\tau} = x e^{-\frac{1}{y}} - \gamma(y - y_0) \tag{8b}$$

where:

$$\beta = \frac{\alpha}{AI^{\frac{1}{2}}}, \quad \gamma = \frac{\alpha}{C\rho AI^{\frac{1}{2}}}$$

It should be pointed out that in the elementary section of the reactor the change in the concentration of the monomer will be insignificant; since total conversion is small, there is a continuous supply of fresh gas and the system is under a constant pressure. On this basis one can assume that the second term of the right-hand side of eqn (7b) is constant

$$\beta(x_0 - x) = m \tag{9}$$

Then, eqn (7b) assumes the form:

$$\frac{dx}{d\tau} = -x e^{-\frac{1}{y}} + m \tag{7c}$$

Now, the models of our chemical system will be represented by eqns (7c) and (8b). Analogous equations were obtained by Salnikov[4] in the investigation of the thermo-kinetic oscillations of chemical reaction $A \rightarrow X \rightarrow B$ for the case of the rate of reaction $A \rightarrow X$ remaining constant.

In order to develop stable periodic solutions (self-oscillations) in the system (7c), (8b) the methods of the qualitative theory of differential equations were used. The study of the non-linear systems of the second order is most expediently carried out by means of a phase plane. The presence of the system of a limiting cycle on the phase plane represents the necessary condition for self-oscillations. In the case here the plane having parameters x and y (concentration of monomer and temperature) is the phase plane. The general procedure of the study is as follows. The states of equilibrium are determined, and the boundary of the region of stable equilibrium states is developed, by means of the equation of the first approximation. After this, using Poincaré's sphere[3], the stability of particular points of the system, in the infinitely remote parts of the phase plane, is determined. If the system has an unstable state of equilibrium and if the infinity is also unstable, then on the basis of Bendixon's theorem[3], it will be possible to arrive at the conclusion that on the phase plane of the system there is bound to be at least one limiting cycle.

By equating the right sides of the eqns (7c) and (8b) to zero

$$-x e^{-\frac{1}{y}} + m = P(x, y) = 0$$

$$x e^{-\frac{1}{y}} - \gamma(y - y_0) = Q(x, y) = 0$$

it is possible to find the equilibrium state coordinates

$$Y_s = Y_0 + \frac{m}{\gamma} \tag{10}$$

$$x_s = m e \frac{\gamma}{y_0 \gamma + m} \tag{11}$$

For the determination of the stability of the equilibrium state we shall introduce new dependent variables

$$x = x_s + \xi, \; y = y_s + \eta$$

and we shall reduce the system (7c), (8b) to two linear equations of the first approximation

$$\frac{d\xi}{d\tau} = a\xi + b\eta, \; \frac{d\eta}{d\tau} = c\varepsilon + d\eta$$

The coefficients of these equations are determined by the following expressions

$$a = p'_x(x_s, y_s), \; b = p'_y(x_s, y_s)$$

$$c = Q'_x(x_s, y_s), \; d = Qy'(x_s, y_s)$$

The necessary and sufficiently conditions of stability of the linear system of the second order are the following equations

$$\sigma = -a - d > 0 \tag{12}$$

$$\Delta = \left| \frac{ab}{cd} \right| > 0 \tag{13}$$

The boundary of the stability region $\sigma = 0$ is determined on the plane m, y_0 by the following equations:

$$m = y_s^2 \left(\gamma + e^{-\frac{1}{y_s}}\right) \tag{14}$$

$$y_0 = y_s \left[1 - y_s \left(1 + \frac{e^{-\frac{1}{y_s}}}{\gamma} \right) \right] \tag{15}$$

The verification of the second condition of stability shows that at any parameters of the system $\Delta > 0$. From this it follows that the equilibrium state is a node or a focus.

For the study of the behaviour of phase trajectories in the infinitely remote parts of the plane G, determinable by the inequalities $x \geqq 0/m$, $y \geqq \varepsilon$, when ε has the smallest desirable positive value, Poincaré's sphere is used. For this, new variables

$$y = \frac{1}{z}, \; x = \frac{\rho}{z}$$

are introduced. Then

$$\frac{d\rho}{d\tau} = zP\left(\frac{P}{z}, \frac{1}{z}\right) - \rho z Q\left(\frac{\rho}{z}, \frac{1}{z}\right)$$

$$\frac{dz}{\sigma\tau} = -z^2 Q\left(\frac{\rho}{z}, \frac{1}{z}\right)$$

where

$$P\left(\frac{\rho}{z}, \frac{1}{z}\right) = -\frac{\rho}{z} e^{-z} + m$$

$$Q\left(\frac{\rho}{z}, \frac{1}{z}\right) = \frac{\rho}{z} e^{-z} - \gamma\left(\frac{1}{z} - y_0\right)$$

Since the identity $P \equiv \rho Q$ does not occur, the equator of Poincaré's sphere ($z = 0$) is an integral curve. The particular points on the equator are determined by the relations $z = 0$ and $P/Q - \rho = 0$. On the equator of the sphere two pairs of particular points $\rho_1 = 0$ and $\rho_2 = 1 + \gamma$ are located.

The subsequent analysis shows that the phase trajectories do not come out of the region G, and on the contour which limits the region, there are no stable states of equilibrium. Therefore, on the basis of Bendixon's theorem[3] it is possible to prove that on the phase plane there is a limiting cycle, which embraces the unstable state of equilibrium.

Thus, the region of unstable states of equilibrium, determinable by eqns (14) and (15), is the region of self-oscillating conditions of the system.

It should be pointed out that if a simplified condition (9) is not adopted, then the study of eqns (7b), (8b) is made difficult by the determination of the state of equilibrium. But the simulation of this system on an analogue computer has shown that in it also, under certain conditions, self-oscillations occur. One of the limiting cycles, obtained on the computer, is represented in *Figure 3*.

In the author's opinion, the self-oscillations of the ethylene polymerization reaction are the main cause of considerable changes in the output of the reactor and quality of the product. Therefore, they should be considered harmful, and it is necessary to search form eans and methods to combat them. This problem is still unsolved.

Automatic Control of the Process

The investigations on the reactor were carried out simultaneously with the automization of the process. The results of investigations were used in solving the problems of automatic control, and the introduction of automatic control has helped the experimental work. Thus, a system of automatic control, the block-diagram of which is shown in *Figure 4*, was constructed. From this diagram it is possible to see which basic functions are performed by this system.

A conventiooal isochromie controller carries out different commands according to pressure changes in the reactor, which are received from other points of the circuit. The controller, in facth, acts as a servo system. After receiving a signal from the pressure-correcting unit, the pressure is gradually reduced if the temperature at any one point of the reactor exceeds the set limit. For the set point of the pressure controller, a signal is also received from the oscillator, which operates on the principle of conventional relay pulse-couple. The oscillator rapidly reduces the pressure in the reactor by 70–100 atm and then gradually raises it to the previous value.

With the appearance of any risk of explosion the safety interlock comes into operation. At first, the pressure in the reactor is reduced, but if this does not result in the prevention of an explosion the contents of the reactor are discharged into the atmosphere. At the same time a signal is sent for the stoppage, of the compressor, and the supply of oxygen is discontinued.

The starting of the reactor is obtained through the command of the operator. The basic operation of starting consists in a gradual increase of pressure in the reactor. If, at the time of starting dangerous operating conditions develop, then the rise in pressure is stopped either automatically or by the command of the operator.

The unloading from the separator takes place periodically through the pressure signal in it. As soon as the pressure in the separator begins to fall, the unloading is stopped, since the reduction in pressure indicates that the separator is completely freed from the liquid polymer. The interval of time between the unloadings is adjusted automatically by a special system, which indirectly measures the output of the reactor and decreases or increases the frequency of unloading. The pressure-control unit in the separator performs simple stabilization of pressure during the intervals between the unloadings. It should be pointed out that the pressure control system in the reactor and that in the separator do not interact.

The unit for the measurement of oxygen provides for the remote automatic (or hand) change in the supply of initiator to the reactor for any programme.

Constructionally, the automatic control system consists of pneumatic control equipment which is designed for the simultaneous automatic control of two reactors. All units of the assembly consist entirely of pneumatic logical components. This provides for adequate reliability and fire risk. A number of such units have been produced and have passed industrial tests at two of the works. Their testing under operating conditions proved their complete reliability and high quality of control.

The proposed system is a natural outcome of only the first stage of work for the automatic control of the process. It embodies the operations which are essential for the maintenance of trouble-free normal operating conditions of the reactor. However the problem of automatic control of the polymerization reaction is not yet completely solved. It may be expected that further study of the process, and particularly of the self-oscillating conditions, will result in the finding of even more efficient methods for the control of the reaction.

Conclusions

As a result of this study a relationship was found between the output of the reactor and the basic parameters of the process. One of the basic causes of the ethylene decomposition was revealed. The self-oscillating conditions in the operation of the reactor were uncovered and the mathematical model of a part of the reactor was studied.

Simultaneously with the investigation of the process, work was carried out for its automatic control as a result of which pneumatic automatic control equipment was constructed.

References

[1] BRANDON, D. B. Developing mathematical models for computer control. *ISA Journal* 6, No. 7 (1959)

[2] HOFTYZER, P. J. and ZWITERING, Th. N. The characteristics of homogenized reactor of the polymerization of ethylene. *2nd Europ. Symp. Chem. Engng.* 1960

[3] ANDRONOV, A. A. Poincaré's limiting cycles and theory of self-oscillations. *Collection of wks of A. A. Andronov, AN SSR*, (1956)41

[4] ANDRONOV, A. A., VITT, A. A. and KHAIKIN, S. E. *Theory of oscillations.* Fizmatgiz (1959)

[5] SALNIKOV, I. E. Theory relating to homogeneous chemical reactions. *Zh. Fiz. khim.* No. 3 (1948)

[6] SEMENOV, N. N. Some problems relating to chemical kinetics and reactive capacity. *AN SSSR* (1958)

[7] FRANK-KAMENTSKII, D. A. Diffusion and heat-transfer in chemical kinetics. *AN SSSR* (1947)

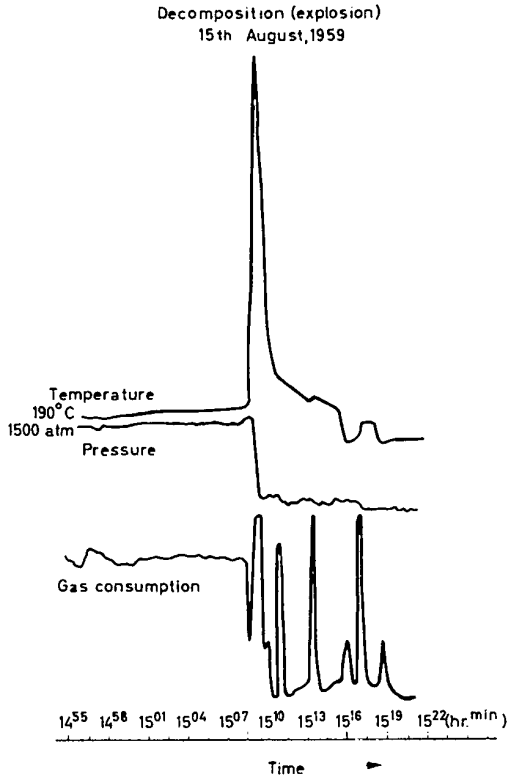


Figure 1. Recording of parameters of the process at the instant of explosion

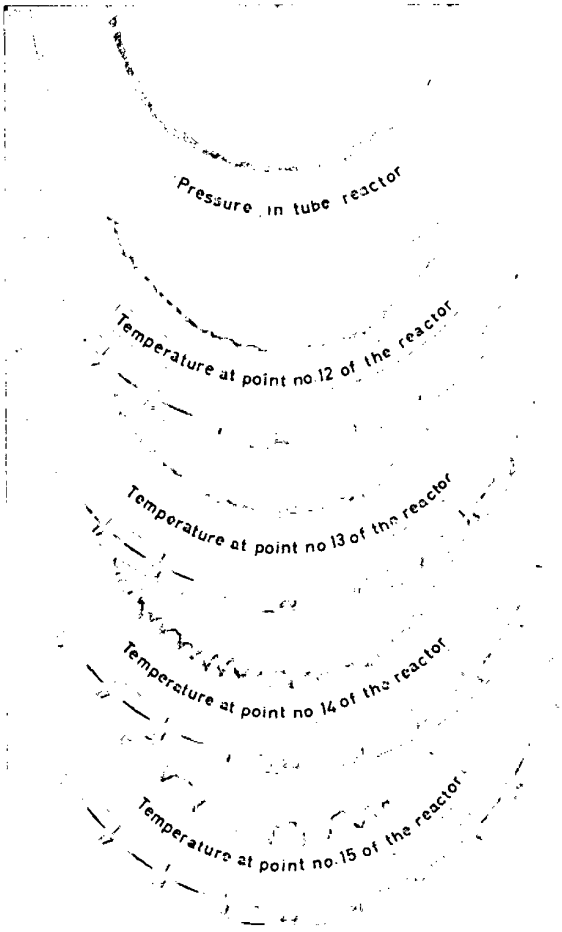


Figure 2. Recording of temperature oscillations

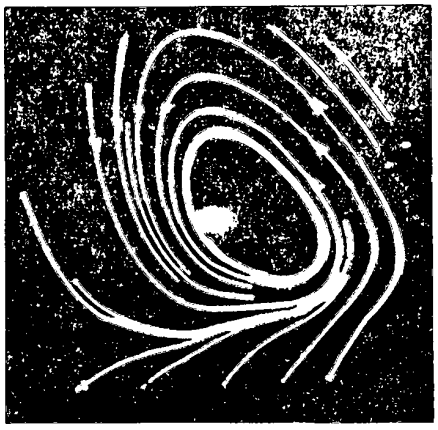


Figure 3. Limiting cycle

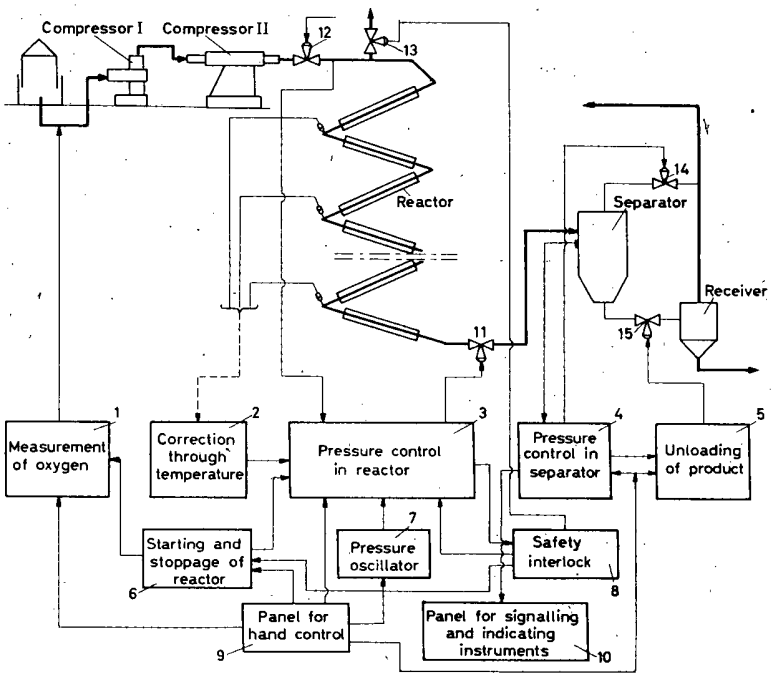


Figure 4. Block diagram of the automatic system of control

A Study of the Dynamic and Static Characteristics of the Process of Fractional Distillation

I. V. ANISIMOV

Introduction

Numerous studies of the dynamics of the process of fractional distillation are based on the consideration of the theoretical and not the actual column plates. For the binary systems the degree of utilization of plates is taken into account but it is assumed that this is independent of the parameters of the process[15–17, 18]. Such a simplified approach introduces substantial errors into the calculations relating to the dynamics and statics of the distillation process.

As a result of studies of the process of fractional distillation for the binary mixtures[1, 3, 8, 9, 11, 12] it was possible to determine the effect of design parameters of the plate, physical and chemical properties of the components and operating parameters of the process on the mass-transfer kinetics. In this work the problems connected with the calculations and analysis of the dynamics and statics of the process for the separation of binary mixtures in the distillation columns are considered in the light of the most recent studies of the mass transfer on the plate, and recommendations are given for the choice of the optimum system of control of the process.

Study of the Dynamic Characteristics of the Process and Special Characteristics which Affect the Choice of the System of Control

A mathematical account of the process was obtained by proceeding from the material balance of the more volatile component of the binary mixture in the distillation column, and the following assumptions were made:

(1) The working of the column is adiabatic.
(2) The liquid is not carried away from the plate.
(3) The mixing within the liquid on the plate and in the vapour is complete.
(4) The quantity of the vapour phase in the column is disregarded.
(5) The pressure on all the plates is equal to that of the atmosphere.
(6) The condenser of the column is full.
(7) All the liquid on the plates is confined to the zone of mass transfer.
(8) The initial mixture and the reflux admitted are at boiling point.
(9) The mass transfer on the column plates is equimolar.
(10) The local mass transfer coefficient at a given instant of time is uniform over the entire plate.

The material balance equations for the more volatile component in the transient process are:

For the top plate

$$H_n \frac{dX_n}{d\tau} = L_D X_D - L_n X_n + V_{n-1} y_{n-1} - V_n y_n \qquad (1)$$

For the feed plate

$$H_f \frac{dX_f}{d\tau} = L_{f+1} X_{f+1} - L_f X_f + V_{f-1} y_{f-1} - V_f y_f + F X_F \qquad (2)$$

For the column still

$$H_0 \frac{dX_0}{d\tau} = L_1 X_1 - V_0 y_0 - W X_W \qquad (3)$$

It is assumed that in the still a single complete evaporation of the liquid portion takes place, under which conditions

$$y_0 = X_0 \qquad (4)$$

In accordance with the assumptions made, the liquid and vapour flow rates are connected by the following equations:

$$V_0 = L_1 - W = V_1 = \ldots = V_W \qquad (5)$$

$$L_D = V_n - D = L_n = \ldots = L_{f+1} \qquad (6)$$

$$L_f = L_{f+1} + F = L_{f-1} = \ldots = L_1 \qquad (7)$$

The formulae, which allow for the hydraulic retardations of the flow, the non-adiabatic character of the process, etc., to be taken into account, are given in another work[2].

For the solution of eqns (1)–(3) it is necessary to determine the relation between the variables.

The assumption about complete mixing of the liquid on the plates makes it possible for the process of mass transfer, which takes place during the motion of a certain volume of the vapour phase through a liquid layer of constant composition, to be considered[14].

The mass-transfer equation for the ith plate may be written in the following form:

$$V_{i-1}\, dy = K_v S_i (y_i^x - y_i)\, d\tau \qquad (8)$$

Assuming that the quantities V_{i-1}, K_v and S_i are constant one obtains

$$y_i = y_{i-1} e^{-\frac{K_{vi}}{V_{i-1}}} + y_i^x \left(1 - e^{-\frac{K_{vi}}{V_{i-1}}}\right) \qquad (9)$$

where

$$K_{vi} = K_v S_i \Delta\tau_i \qquad (10)$$

The general mass transfer coefficient on the plate K_{vi}, determinable by plate design, physical and chemical properties of the components and by operating parameters, makes it possible for the effect of these factors on the transient process to be taken into account in the calculations.

According to the double resistance theory[13], the general mass transfer coefficient is a function of the particular mass transfer coefficients of the liquid and vapour phases:

$$K_{vi}=\frac{1}{\frac{1}{\beta vl}+k_i\frac{1}{\beta li}} \qquad (11)$$

where

$$k_i=\left(\frac{\partial y^x}{\partial x}\right)_i$$

the phase equilibrium constant.

The particular mass transfer coefficients may be calculated on the basis of experimental data as definite functions of the plate design parameters, physical and chemical properties of the components, composition of the liquid and vapour phases on the plate and of vapour or liquid flow rates in the column[10].

The system of eqns (1)–(11) describes the transient process in the fractional distillation column for the separation of binary mixtures, taking into account the kinetics of mass transfer on the plates.

As an example, the calculation and the analysis of the transient processes for the separation of the methanol–water mixture in a distillation column are given. The initial data are as follows: the pressure in the column is atmospheric; the number of plates $n = 18$; the feed plate number $f = 9$; the quantity of still product $W = 166{\cdot}5$ kg-mole/h; the quantity of initial mixture $F = 229{\cdot}2$ kg-mole/h; the quantity of distillate $D = 62{\cdot}7$ kg-mole/h; the quantity of vapour $V_0 = 141{\cdot}1$ kg-mole/h; the concentration of the more volatile component in the feed $X_F = 0{\cdot}273$ mole fractions; the concentration of the more volatile component in the distillate $X_{19} = 0{\cdot}973$ mole fractions; the concentration of the more volatile component in the still $X_0 = 0{\cdot}0085$ mole fractions.

$$\beta_{vi}=1{\cdot}61\,V_{i-1}\,46\,\text{kg-mole/h/plate surface.}$$

$$\beta_{li}=380\,\text{kg-mole/h/plate surface.}$$

The calculations for the transient processes in the column were carried out on a universal digital computer for the following step-like unit disturbances:

(1) For an increase in the concentration of the more volatile component of the initial mixture

$$\Delta X_F=X_F\times\frac{5}{100}$$

(2) For an increase in the quantity of feed

$$\Delta F=F\times\frac{5}{100}$$

(3) For an increase in the distillate withdrawal

$$\Delta D=D\times\frac{5}{100}$$

(4) For an increase in the quantity of vapour leaving the evaporator

$$\Delta V_0=V_0\times\frac{5}{100}$$

The calculation results are given in the form of response curves in *Figures 1–4*. The curves obtained by calculations based on theoretical plates are shown by dots. The comparison of curves shows that the results of calculations based on the theoretical plates and those based on the proposed method are substantially different, especially for the plates of the low separating capacity.

By comparing the response curves it is possible to record the following basic dynamic characteristics of the fractional distillation process, which affect the choice of the control system:

(1) The greatest effect on the transient processes and on the concentration distribution along the column height in the state of equilibrium is shown by disturbances which violate the conditions of the material balance in the column, especially by those connected with a change in the distillate withdrawal.

(2) The transient processes in the column take place slowly; in the example considered they require from 1·7 to 2·5 h. The response time of the column depends on the number of plates, relative volatility of the components and other factors[2].

(3) The changes in the concentration of the liquid on the upper and lower plates of the column are insignificant. The greatest changes in the concentration of the liquid take place in the so-called 'controlling' plates, which are situated approximately in the middle of the evaporating and restorative sections of the column. The position of the 'controlling' plates may be considered independent of the form of disturbances.

The input selection for the control of composition or temperature of the liquid should be made from one of the controlling plates. On no account is it possible to control the process directly through the composition of distillate or still product, since the static and dynamic characteristics of the process would deteriorate rapidly.

(4) The change in the steam supplied to the evaporator gives rise to transient processes in the draining and restorative sections of the column, which are different in character. This is attributed to the action of two opposing factors: to an increase in the separating capacity of the column with the increase in the reflux number, and to a decrease in the efficiency of each plate with an increase in the vapour flow rate. At the very beginning the changes in concentration for the restorative and draining sections of the column have different signs.

(5) In a transient process considerable delays in the change of composition (of temperature) of the liquid phase occur. The delays in the change of composition of the vapour phase on the plates caused by the change in the vapour flow rate in the column are considerably smaller. This is explained by the fact that the value V of the vapour flow changes with a speed which is close to that of sound; therefore, the conditions of mass transfer on the plates change almost instantaneously, see eqn (9). This phenomenon finds no explanation in calculations based on the theoretical plates.

In the overwhelming majority of cases the control circuits for the process of fractional distillation are limited to the prob-

lem of stabilization of the parameters of the process[2]. Such automatic control systems work more or less satisfactorily if the disturbances are small and if variations in the quality of the product are permissible. With appreciable changes in the quality and composition of the initial mixture the continuous deviations from the assigned composition of distillate and still product are unavoidable. In order to obtain products of high purity under these conditions the invariance of the process control systems is the most desirable.

A system of control cannot be made absolutely invariant in respect of all the disturbances. In the fractional distillation process the violations of the material balance caused by changes in the quality and composition of the initial mixture represent the basic disturbances. The violations of the thermal balance of the process, the changes in pressure in the column, the variations in the quantity of liquid on the plates and in the still, the changes in the working efficiency of the plates caused by change in the composition of the feed and in the vapour flow rate in the column, etc. represent the less important and secondary disturbances.

It is possible and expedient to construct a selective invariant system of control, for which the basic parameter of the process—the composition of the liquid on the control plate—will be independent of the changes in the quantity and composition of the initial mixture.

With a selective invariant system of control only small changes in the composition of the liquid on the control plate under the action of the less important secondary disturbances of the process will occur. Therefore, the system of control should be based on the combination of principles of control according to disturbance and deviation of parameter.

An account of the fundamentals of the theory of combined control and of the condition of invariance are given in other works[4–7].

The amplitude and phase characteristics of the controlled plant according to control and disturbance paths required for the calculation of the conditions of invariance, are not difficult to determine from the response curves obtained as a result of the solution of the system of equations for the dynamics of the process.

The changes in the quantity and composition of the initial mixture violate simultaneously the material and the thermal balance of the process. The system of control, which reacts to these disturbances, compensates for their effect in the column by the corresponding change in the supply of the reflux and heating vapour. The oscillations in the pressure of the heating vapour and reflux and the inaccurate readjustment of the control elements represent the secondary disturbances, the effect of which may be easily eliminated by applying flow ratio controllers, which measure the magnitude of disturbance and of response change in the supply of the controlling means.

The selective invariant system of control does not embrace the controllable parameters, which have a smaller effect on the dynamic and static characteristics of the process. These parameters are stabilized by customary controllers.

On the basis of what has been stated, a block diagram for a combined selective invariant system of control for the process of fractional distillation (described at the end of this paper—see *Figure 7*), has been developed.

The Static Characteristics of the Process

The task of automatic control consists in the determination and maintenance of the optimum values of the controlling parameters of the process.

The calculated values of the following parameters of the fractional distillation process are considered to remain approximately unaltered under operating conditions: the pressure in the column, the level of the liquid in the still of the column, the level of reflux, and the temperature of the initial mixture and reflux. The control of these does not present any difficulties and is not shown in the diagram of *Figure 7*.

The optimum values for the reflux number, the quantity of the heating vapour and the location of feed plate change under operating conditions. In the separation of multi-component mixtures it is necessary to determine also the optimum quantities and points of withdrawal for the intermediate products.

The optimum values of these parameters based on the minimum cost of manufacture are determined as the functions of the quantity and composition of the initial mixture, provided that the product obtained is of precisely the composition assigned or that it changes within the permissible limits.

For the calculations relating to the statics of the fractional distillation process, the material balance equation for the state established in the part of the column situated below the i–lth plate is written

$$L_{i+1}X_{i+1}-V_i y_i+FX_F-WX_0=0 \qquad (12)$$

where

$$L_{i+1}=V+W \text{ when } i<f \text{ and } L_{i+1}=V_i+W-F \text{ when } i\geq f \qquad (13)$$

$$V_i=V \text{ when } 0\leq i\leq W \qquad (14)$$

Consequently, the material balance of the process for the established state may be written in the form:

$$X_i=\frac{1}{V+W}(Vy_{i-1}+WX_0) \text{ when } 0\leq i<f \qquad (15)$$

$$X_i=\frac{1}{V+W-F}(Vy_{i-1}+WX_0-FX_F) \text{ when } f<i\leq n+1 \qquad (16)$$

The statics of the fractional distillation process is described by the system of eqns (4), (8)–(11), (15) and (16). Its solution makes it possible to obtain the static relations between the basic parameters of the process and the concentration distribution of the more volatile component in the liquid on the plates for different operating conditions.

The calculation of the static characteristics of the process was made for the above-mentioned fractional distillation column for the separation of the methanol–water mixture, for the different quantities and compositions of the initial mixture, and for the constant composition of distillate and still product. As an example, in *Figures 5* and *6* the static characteristics of the column are given. From *Figure 5* it is evident that within a certain range of values for the concentrations X_F and loads G_F there exists an extremum relationship for the steam consumption Q per unit weight of distillate G_D. With the increase in G_F the heat consumption per unit of G_D also increases, especially at high concentrations X_F. From the graph it is possible to determine

the operating conditions for which the energy requirements will be within the limits which are economically expedient.

From the consideration of *Figure 6* it follows that the static characteristics have an extremum and ambiguous values (the assigned compositions of the final products may be obtained under different operating conditions). Curves I and II, which limit the operating region for the parameters of the process, represent the locus of values of the coordinates V and D, at which the compositions of the final products are exactly equal to those assigned. The minimum energy requirements of the process correspond to the minimum value for the vapour flow V which, at the given values of D, F and X_F will secure the assigned compositions X_D and X_W. One of the tasks of the optimum control is the determination and the maintenance, in relation to the values of F and X_F, of the values V and D, which correspond to the coordinates of points situated on the left side of the static characteristics.

For each set of operating conditions there is a limiting load for the column in respect of the quantity of the initial mixture of a given composition, at which the operating region degenerates into a point, see the extremum on curve I. With a further increase in the quantity of the initial mixture it is impossible to obtain the assigned compositions for the final products.

A reduction in load decreases the necessary vapour flow, which leads to an increase in the enrichment of the vapour phase by the more volatile component, and to an increase in the efficiency of mass transfer, see eqn (9).

The optimum place for the introduction of the initial mixture into the column is determined for each set of operating conditions, proceeding from the fact that the concentration of the more volatile component in the initial mixture X_F should be equal to the concentration X_f on the feed plate, i.e., the following condition is observed:

$$X_{f-1} < X_F < X_{f-1}. \qquad (17)$$

As a result of the analysis of calculations relating to the statics of the process it is possible to make the following deductions:

(1) The plate-type distillation column for the separation of binary mixtures is a non-linear system. The independent parameters in the calculations relating to the statics of the process are the load of the column based on the quantity of the initial mixture F, the composition of the initial mixture X_F, the value of the vapour flow rate in the column V and the distillate withdrawal rate D.

(2) The region of the static characteristics in which the conditional products may be obtained is limited by the four independent parameters indicated. These limitations are conditioned by the kinetics of mass transfer. The assignment of values for X_D and X_W, which fall outside the region of their joint existence, may cause oscillating operating conditions in the column (the conditions of joint existence of values for X_D X_W are realized periodically).

(3) The relation between the final products of the column and the vapour flow rate may have an extremum. An increase in the vapour flow rate increases the motive force of the process $Y_i^x - Y_i$, but reduces the efficiency of each plate, which gives rise to the extremum. This phenomenon is not found in the calculations based on theoretical plates. The extremum for the static characteristics may be conditioned by the kinetics of mass transfer, as well as by the carrying away of the liquid from the plates.

(4) The static characteristics are ambiguous. This property develops only in calculations which take into account the kinetics of mass transfer on the plates. The range of characteristics, situated on the left side of the extremum, represents the operating range.

(5) The change in composition of the vapour phase on the plates is usually more appreciable than that for the liquid phase.

(6) The optimization of the process produces increased demands on the system of automatic control, in view of the steepness and ambiguity of the static characteristics.

As a result of the investigations described it was possible to develop a control system for the distillation process, which is shown in *Figure 7*. Controller 1 maintains the assigned optimum rate of supply of the initial mixture to the column.

Instruments 4 and 5 measure the rate of flow of the initial mixture and send signals to controllers 2 and 3 for the flow ratios G_F/G_R and G_F/Q.

The dynamic characteristics of instruments 4 and 5 are computed so that the conditions of selective invariance in respect of disturbances for the rate of flow of the initial mixture are fulfilled. Controllers 2 and 3 maintain the material and thermal balance of the process.

The control based on the disturbance of composition of the initial mixture and on the deviation of the composition of the liquid on the controlling plate is achieved by these same controllers through the assignments computed and set by computer 10 (universal digital computer).

Converters 7 and 8 receive signals from transducers 6, 9 and 11 which measure the compositions X_F and X_i and the rate of flow G_F, and transform them into signals which in turn are admitted to computer 10.

The computer performs the following operations:

(1) Calculation of the optimum load of the column G_F for the current values of X_F and setting of the assignment for the rate of flow controller 1, see *Figure 5*.

(2) Calculation of optimum ratios G_F/G_R and G_F/Q in relation to the current values of G_F and X_F and setting of the assignment for controllers 2 and 3, conforming to the conditions of selective invariance.

(3) Correction of the calculated optimum ratios G_F/G_R and G_F/Q based on the degree of deviation of the basic controllable parameter—the deviation of concentration of the more volatile component in the liquid on the selected plate (closing of the control loop by means of the feed-back signal).

(4) Calculation of the optimum feed plate number and shifting of the inlet of the initial mixture to the necessary plate.

(5) In the case of multi-component mixtures: calculation of the plate number for the withdrawal of the side product and calculation of its quantity. The corresponding assigned operations: the changing over to the necessary withdrawal plate and setting of the assignment for the controller of the side product flow rate are not shown in the diagram.

(6) Transition from one algorithm of control to another—in accordance with the change in the optimization assignment, with

the transition (having reached definite parameters values) from starting to normal operating conditions and from the latter to the shut-down, etc.

In addition to this the usual operations of automatically checking the accuracy of calculations and the working order of the computer, the printing of results, signalling of inaccuracy and faults, etc. should be performed. In case of faults or stoppage of the computer, the assignments to controllers should remain at values determined at the preceding instant.

In the development of the considered control circuit it was assumed that the temperature of the initial mixture is constant. It is known that the heating of the mixture to its boiling point represents the optimum condition. With the variable composition and constant temperature of the initial mixture the ratio between the liquid and the vapour phase, and the enthalpy will change. Therefore, in the case of the composition of the initial mixture changing within wide limits, it is expedient to control its enthalpy. For this, an instrument should be included in the control circuit which would measure the enthalpy of the initial mixture and send the signal to the computer. The computer should calculate the optimum enthalpy value for the parameters of the initial mixture at the corresponding instant of time and pass the assignment to the steam consumption controller, which in turn should transmit it to the heat exchanger for feed heating.

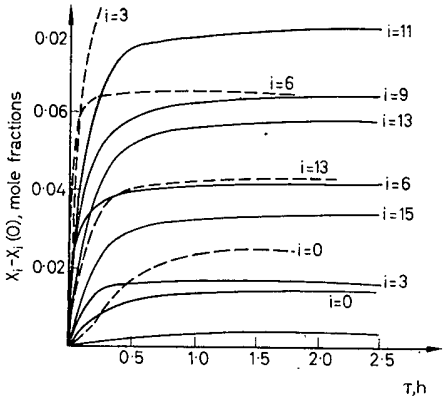
The adaptation of the proposed control system is expedient in those complex cases where it is required that the separation of components of the mixture should be made with a high accuracy and where optimization of the process is required.

Nomenclature

D Quantity of distillate (kg-mole/h)
W Quantity of still product (kg-mole/h)
F Quantity of initial mixture (kg-mole/h)
i Plate number, for still $i = 0$, for condenser $i = n + 1$
f Feed plate number
H Quantity of liquid on the plate (kg-mole)
L Quantity of liquid running off the plate (kg-mole/h)
V Quantity of vapour leaving the plate (kg-mole/h)
Q Quantity of heat supplied to the evaporator (kcal/h)
$G_{F,V,D,W,R}$ Quantity of initial mixture, vapour, distillate, still product, reflux (kg/h)
x Concentration of the more volatile component in the liquid on the plate (mole fractions)
y Concentration of the more volatile component in the vapour above the plate (mole fractions)
y^x Concentration of the more volatile component in the vapour which is in a state of equilibrium with the liquid of composition x (mole fractions)
K_v General mass-transfer coefficient, related to the unit area of phase contact, calculated by the vapour phase (kg-mole/m²/h)
S Phase contact area on the plate (m²)
β_1 Particular mass-transfer coefficient in the liquid phase (kg-mole/m²/h)
β_v Particular mass-transfer coefficient in the vapour phase (kg-mole/m²/h)
τ Time (h)
$\Delta\tau$ Contact time of phases on the plate (h)

References

[1] AKSELROD, D. S. *Doctorate Thesis.* MIKhM (1958)
[2] ANISIMOV, I. V. *Automatic Control of Fractional Distillation Columns.* 1961, 2nd Edition, Gostoptekhizdat
[3] DILMAN, V. V., OLEVSKII, V. M. and KOCHERGIN, N. A. Theory and practice of fractional distillation in the chemical and food industries. *All-Union Inter-Inst. Conf.* (Collection of Reports). *Izdatel'stvo Kievskogo Universiteta* (1960)
[4] IVAKHENKO, A. G. Tekhn. Kibern., Kiev (1959)
[5] KULEBAKIN, V. S. Methods of improving the quality of automatically controlled systems. *Proc. Zhukovskii's Inst.*, Ex. 5021 (1954) 3–51. Trudy VVIA im. Zhukovskogo
[6] KULEBAKIN, V. S. Plate-type fractional distillation. *Proc. 2nd All-Union Conf.* Izdatel'stvo AN SSSR, Vol. P (1955) 184–207
[7] KULEBAKIN, V. S. *Proc. Sov. Acad. Sci.*, Vol. 68, No. 5 (1949); Vol. 77, No. 2 (1951)
[8] KASATKIN, A. G., PLANOVSKII, A. N. and CHEKOV, O. S. *Calculations relating to fractional distillation and absorption equipment.* 1961, Standartgiz
[9] ORLOV, B. N. *Thesis.* MKhTI (1961)
[10] PLANOVSKII, A. N. and NIKOLAEV, P. I. *Processes and Equipment of the Chemical and Petroleum Industries.* 1960. Gostoptekhizdat
[11] SOLOMAKHA, V. P. *Thesis,* MIKhM (1957)
[12] CHEKHOV, O. S. *Thesis,* MIKhM (1959)
[13] LEWIS and WHITMAN *Industr. Engng. Chem. (Industr.)* 16 (1924) 125
[14] MURPHREE *Industr. Engng. Chem. (Industr.)* 17 (1925) 747
[15] ROSENBROCK, . Calculation of the transient behaviour of distillation columns. *Brit. Chem. Engng.*, 3 (1958) 363–367, 432–435, 491–494
[16] RADEMAKER, O. and RIJNSDORG, J. E. Dynamics and control of continuous distillation. *5th World Petroleum Congr.* London (1960)
[17] WILKINSON, W. L. and ARMSTRONG, W. D. *Plant and Process: Dynamic Characteristics,* 56–72, London (1957)
[18] VOETTER, H. *Plant and Process: Dynamic Characteristics,* 73–100, London



← *Figure 1. Response curves for concentrations X_i for a step-like unit increase in X_F amounting to 5 per cent*

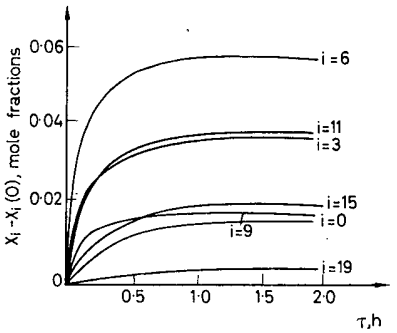


Figure 2. Response curves for concentrations X_i for a step-like anit increase in the quantity of the initial mixture amounting to 5 per cent

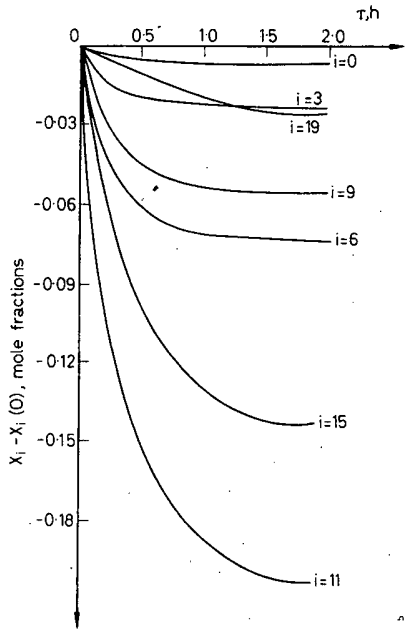


Figure 3. Response curves for concentrations X_i obtained for a step-like unit increase in distillate withdrawal amounting to 5 per cent

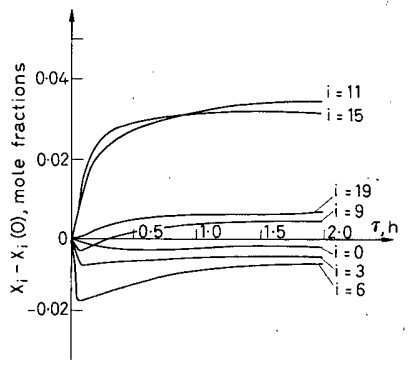


Figure 4. Response curves for concentrations obtained for a step-like unit increase in vapour flow rate in the column amounting to 5 per cent

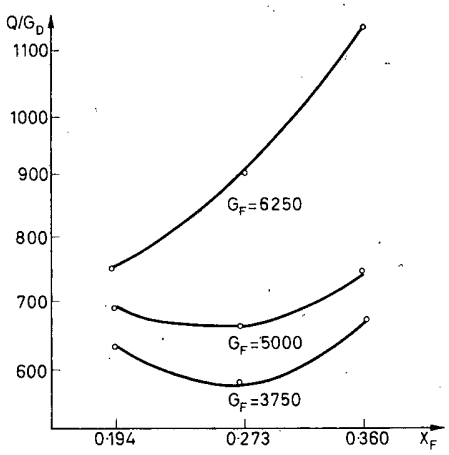


Figure 5. The graph illustrating the relationship between the heat consumption per unit weight of the distillate Q/G_D, and the quantity G_F and composition X_F of the initial mixture

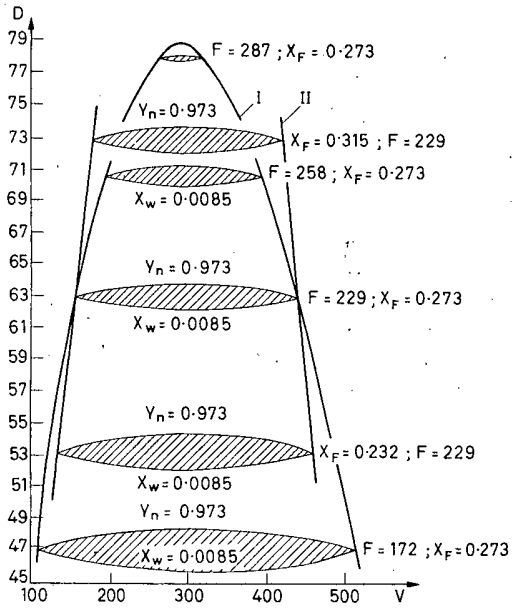


Figure 6. The operating region for the static parameters of the fractional distillation column. Curve I: quantity of the initial mixture F is variable, whilst its composition X_F is constant. Curve II: composition of the initial mixture is variable, whilst its quantity is constant

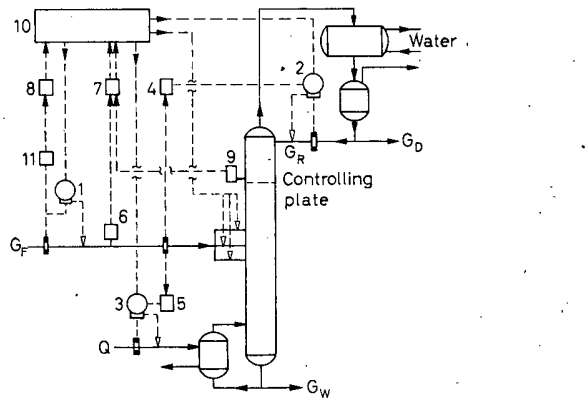


Figure 7. Block diagram for a combined selective invariant system of control for the process of fractional distillation

The Realization of a Self-adapting Control Programme in a System with a Digital Computer

P. F. KLUBNIKIN

Introduction

Recently there has been a wide expansion in systems in which the control of a load is achieved using a digital computer. In these systems by means of the application of the proper control system it is possible to obtain a self-adaptive (self-organizing) property, even in the case where there is no *a priori* information on all the properties of the load and the change of the load parameters in the course of time.

The elements of the theory of the construction of self-adaptive control systems are known, but the application of them in practice often results in substantial difficulties connected with the special digital computing systems. The main difficulty is the determination of the characteristics of the load (for example, the transfer function) under conditions of normal operation of the system and the search for the control signal input to the load, which gives, from one or another point of view, the best control process.

This paper is devoted to the questions of the realization of a self-adaptive control programme in a system with a digital computer. One method of constructing a self-adaptive control programme is considered, which permits it to be realized relatively simply. Results of the experimental investigation of control systems are presented.

The Method of Self-adaptation in the Control Programme

Consider an automatic control system consisting of a continuous part (the load) and a digital computer (DCM, *Figure 1*). The DCM operates in a realm of periodic repetition of a programme with a time cycle T.

Let the link between the control input X_0 and the output quantity of the system X be given in the form

$$X^* = W_3(z)\, X_0^* \tag{1}$$

where X^* and X_0^* are the values of X and X_0 at the moment of time T; $W_3(z)$ is the transfer function of the instantaneous system; $z = e^{-q} = e^{-Ts}$ is a lag operator.

Then, as is known[1–3], in order for (1) to be satisfied the control system can be realized in the form of the block diagram shown in *Figure 2*, where for the transfer functions of the elements of the programme the following conditions should be satisfied

$$D_1(z) = W_3(z), D_3(z) = \frac{W_3(z)}{W_H(z)} \tag{2}$$

where $W_H(z)$ is the discrete transfer function of the load. The transfer function $D_2(z)$ is chosen arbitrarily. In the most simple case

$$D_2(z) = \frac{C_1 z + C_2}{C_3 z + 1} \tag{3}$$

However, to satisfy the conditions (2) and, consequently to obtain the prescribed properties of the system, is impossible when the transfer function of the load $W_H(z)$ is unknown or when its coefficients have an unknown time dependence, which is often the case in practice. It should be noted, that in the indicated situation a general control programme, calculated in the presence of complete information about the load, is usually not convenient.

Therefore, the first step in a self-adaptive control programme is the determination of the discrete transfer function of the load, which is written in the form

$$\frac{X^*}{d^*} = W_H(z) = \frac{A_n z^n + A_{n-1} z^{n-1} + \dots A_1 z}{B_n z^n + B_{n-1} z^{n-1} + \dots B_1 z + B_0} \tag{4}$$

where n is the order of the load equations; $A_i(t)$, $B_j(t)$ are time-dependent coefficients ($i = 1, 2, 3, \dots, n$, $j = 0, 1, 2, \dots, n$).

Consider that the computing-time cycle of the DCM is chosen so that the coefficients $A_i(t)$ and $B_j(t)$ are unable to change significantly over several cycles, and that n is unknown. Then in order to determine during the process of operation of the system the current values of the coefficients A_i and B_j, and consequently $W_H(z)$ for a given moment of time, it is possible to use two simpler methods.

The first method is similar to that described in a previous work[4] and is based on the solution of a system of equations, which is obtained by using the expressions (4), i. e.

$$X_k B_0' + X_{k+1} B_1' + X_{k+2} B_2' + \dots + X_{k+n} B_n' = d_{k+1} A_1' + d_{k+2} A_2' + \dots + d_{k+n} A_n' \tag{5}$$

where $k = 0, 1, 2, \dots, 2n$

$$d_k = d(t - kT)\, u X_k = X(t - kT)$$

are the values of the input and output quantities of the load measured in the kth preceding calculation cycle; A_i' and B_j' are the approximate values of the coefficients for the current calculation cycle.

The system of eqns (5) is solved on the DCM relative to A_i' and B_j' by one of the known methods, for example by the method of iterations, and $W_H(z)$ is determined in the same way. The second method uses the principle of a 'learning model'[5] and includes the following.

Using the values of X_k and d_k ($k = 0, 1, 2, \dots, n$) available in the memory of the DCM, a search is carried out by the

gradient method for the magnitudes of the coefficients A_i, B_j, which give a minimum in the mean difference

$$\Delta_{av}=\frac{1}{\overline{m}}\sum_{v=0}^{m}|X(t-vT)-X_m(t-vT)|$$

where $\overline{m}$ is the number of cycles for averaging

$$X^*=W_H(z)\,d^*$$

$$X_m^*=W_{HM}(z)\,d^*$$

$W_{HM}(z)$ is a 'model' transfer function for the load formed in the DCM. This method is illustrated in the diagram shown in *Figure 3*.

The second stage of the method described for building a self-adaptive control programme is the determination of a control signal d, which will guarantee the stability of the system and satisfy (1) or a better approximation to this condition. As a criterion for the approximation to (1) it is more useful to select the mean absolute value or the mean square of the error

$$\varepsilon_{av}=\frac{1}{\lambda}\sum_{v=0}^{\lambda}|\varepsilon(t-vT)| \qquad (6)$$

$$\varepsilon_{av}=\frac{1}{\lambda}\sum_{v=0}^{\lambda}[\varepsilon(t-vT)]^2$$

where λ is the number of averaging cycles:

$$\varepsilon^*=[W_3(z)-W_3'(z)]\,X_0^*$$

$W_3'(z)$ is the transfer function of the instantaneous system,

$$W_3'(z)=\frac{W_H(z)[D_1(z)\,D_2(z)+D_3(z)]}{1+D_2(z)\,W_H(z)} \qquad (7)$$

Substituting in (7) the values of $D_1(z)$ and $D_2(z)$ from (2), one gets

$$W_3'(z)=\frac{\dfrac{W_H(z)}{W_{HM}(z)}+D_2(z)\,W_H(z)}{1+D_2(z)\,W_H(z)}\,W_3(z)$$

Obviously in the general case $W_{HM}(z)\neq W_H(z)$ and consequently $W_3'(z)\neq W_3(z)$. However, as experimental investigations have shown, even a relatively rough approximation $W_{HM}(z)$ to $W_H(z)$ for the condition of stability of the instantaneous circuit of the system (*Figure 2*), gives a behaviour of the system that is close to that prescribed.

Thus on the basis of (6) and (7) one has

$$\varepsilon_{av}=F(C_1,C_2,C_3,\ldots) \qquad (8)$$

The stability of the system reaches that sought in the region of the coefficients C_1, C_2, and C_3 of the minimum ε_{av}. The search is carried out by means of extrapolation in the DCM of X_0 in r-conditional cycles and the calculation of ε_{avE} for these cycles.

The idea of the method is explained in the diagrams shown in *Figure 4*. As a result, for each cycle of the DCM the following order of operations is obtained:

(1) *Input X_0 and X.*

(2) *Extrapolation of X_0 in r-conditional cycles.*

In the simplest case for linear extrapolation from the preceding cycle one gets

$$X_{0E}(t+zT)=z[X_0(t)-X_0(t-T)]+X_0(t) \qquad (9)$$

(3) *Determination of the coefficients of $W_H(z)$.*

(4) *The search for the minimum ε_{avE} in the region of the coefficients $D_2(z)$ taking into account the next r cycles.*

For the method of the modified gradient, on the basis of (8), one has the following formulae

$$\varepsilon_{avE}^{(C_1)}=\frac{F_\Delta(C_1+\Delta C,C_2,C_3)-F(C_1,C_2,C_3)}{\Delta C}$$

$$\varepsilon_{avE}^{(C_2)}=\frac{F_\Delta(C_1,C_2+\Delta C,C_3)-F(C_1,C_2,C_3)}{\Delta C}$$

$$\varepsilon_{avE}^{(C_3)}=\frac{F_\Delta(C_1,C_2,C_3+\Delta C)-F(C_1,C_2,C_3)}{\Delta C}$$

$$\Delta C_1=-k\varepsilon_{avE}^{(C_1)}$$

$$\Delta C_2=-k\varepsilon_{avE}^{(C_2)}$$

$$\Delta C_3=-k\varepsilon_{avE}^{(C_3)}$$

$$D_2'(z)=\frac{(C_1+\Delta C_1)z+C_2+\Delta C_2}{(C_3+\Delta C_3)z+1}$$

$$X_{1E}^*=W_3(z)\,X_{0E}^*,\quad X_{2E}^*=(X_{1E}^*-X_E^*)\,D_2'(z)$$

$$X_{3E}^*=\frac{W_3(z)}{W_{HM}(z)}X_{0E}^*,\quad X_E^*=W_{HM}(z)\,d_E^*$$

$$d_E^*=X_{2E}^*+X_{3E}^* \qquad (10)$$

[*Note:* the quantity ΔC can be taken equal to unity.]

where $\varepsilon_{avE}^{(c)}$ are the partial derivatives with respect to the coefficients $D_2(z)$; k is the coefficient of a step in the direction of the reversed gradient; ΔC is the trial increment; X_E, X_{1E}, X_{2E}, X_{3E}, d_E are the values of the corresponding quantities in the conditional cycles within the DCM (the index E indicates extrapolated values of the corresponding quantities).

(5) *After an m step search the output signal from the DCM d is calculated in accordance with the diagram shown in Figure 2.*

$$d(t)=X_2(t)+X_3(t)$$

$$X_2(t)=C_2[X_1(t)-X(t)]+C_1[X_1(t-T)-X(t-T)]$$
$$-C_3X_2(t-T)$$

$$X_1(t)=G_1X_0(t-T)+G_2X_0(t-2T)+\ldots+G_mX_0(t-mT)$$

$$X_3(t)=\frac{1}{A_1}G_1X_0(t)+\frac{1}{A_1}(G_1B_1+G_2)X_0(t-T)$$

$$+\ldots+\frac{1}{A_1}B_nG_mX_0[t-(m+n-1)T]-\frac{A_2}{A_1}X_3(t-T)$$

$$-\ldots-\frac{A_n}{A_1}X_3[t-(n-1)T] \qquad (11)$$

Here one takes

$$W_3(z)=G_1z+G_2z^2+G_3z^3+\ldots+G_mz^m$$

(6) *The output of the control signal d and the updating of the information in memory.*

In *Figure 5* is shown the logical flow diagram of a self-adaptive DCM programme which assures that the given operations and the calculations according to the formulae (9)–(11) will be carried out. Circles indicate conditional transfer operators (transfer control), and the conditions are written inside them. As can be seen from the flow chart, 15 cycles are provided for in the control programme, in the course of which normal control is achieved according to the diagram shown in *Figure 2*. This is required for the accumulation of information in the DCM. No particular explanation is required for the remainder of the flow chart.

It must only be noted that the number of conditional cycles in the DCM must be chosen so that the time for accomplishing the operations described does not take longer than the cycle time T. If, for a minimum number of conditional cycles (one or two), it is not possible to satisfy this condition, then it is necessary to use a DCM that is faster acting (in which each arithmetical or logical operation is executed in less time).

Results of Experimental Investigations

In carrying out the experimental investigation of the load in the system of *Figure 1* its dynamic model was changed. The dynamic model of the load was linked with the DCM through a device transforming a voltage into an 8-digit binary code or the code into a voltage. The control input X_0 is supplied in the form of a voltage and fed through the transforming device to the DCM. The diagram for the realization of the control system during the performance of the experiment is shown in *Figure 6*.

A control programme was fed into the DCM corresponding to the flow diagram shown in *Figure 5*. The dynamic model of the load was characterized by the transfer function

$$W_H(S)=\frac{K(T_0S+1)}{S(T_1^2S^2+2T_1\xi S+1)} \qquad (12)$$

The quantities K, T_0, T, and ξ can be varied in time over the following limits: $K = 0{\cdot}1 \div 0{\cdot}001$; $T_0 = 1{\cdot}5 \div 0{\cdot}2$; $T_1 = 0{\cdot}2$–$0{\cdot}5$; $\xi = 0{\cdot}2 \div 0{\cdot}05$.

The rate of change of the quantities indicated did not exceed 1–5 per cent/sec from the initial value. The calculation cycle in the DCM was equal to $T \cong 0{\cdot}15$ sec. The connection between the control input X_0 and the output of the system X was given in the form

$$W_3(z)=\frac{1}{3}(z+z^2+z^3) \qquad (13)$$

For fixed values of the coefficients $W_H(s)$ of (12) and with fulfilment of the conditions (2) the system has a first-order instability and a transfer process defined by (13). The rate gain in the system is relatively small. Its increase is limited by an instability in the instantaneous circuit for the selected structure $D_2(z)$ of (3).

In *Figure 7* is shown an oscillogram for the development of a control system with normal control (self-adaptive programme excluded). As the experiment shows for normal control the system is extremely sensitive to a change of the coefficients $D_2(z)$, especially when this leads to an increase in the gain of the instantaneous circuit. In this case a change in the coefficients C_1, C_2, and C_3 by 10–15 per cent makes the system unstable. The same effect occurs in the system with a change in $W_H(z)$.

On putting a self-adaptive control system into operation for a short time (10–15 sec) the optimal value of the coefficients $D_2(z)$ was found and the error was reduced to a minimum. In the process of operating, the system automatically adapted itself to the changed characteristics of the load.

In *Figure 8* are shown typical curves of the change of the coefficients $W_H(z)$ during their determination in the DCM. The transfer function corresponding to (12) is written in the form

$$W_H(z)=\frac{a_1z+a_2z^2+a_3z^3}{(1-z)(a_4z^2+a_5z+1)} \qquad (14)$$

It can be seen from the curves that even for 6–8 sec the coefficients of (14) a_i approximate their true values; indicated on the graph by broken lines.

The curves in *Figure 8* were made for the very worst case, where the determination is carried out by a step input to the system applied at the time $t = 0$, after which X_0 remains constant. For an arbitrary time change of $X_0(t)$ the errors in the determination of the coefficients are significantly decreased and do not exceed 5–10 per cent.

In *Figures 9* and *10* are shown oscillograms showing the evolution of the system in the process of changing the coefficients $D_2(z)$, C_1, C_2, and C_3. The oscillograms in *Figure 9* correspond to a combination of initial values of C_1, C_2, and C_3 for which the total gain *pa* of the instantaneous circuit of the system, i.e., $W_H(1)\,D_2(1)$ is small and for which the variable input signal $X_0(t)$ error is large. The oscillograms in *Figure 10* correspond to initial values C_1, C_2, and C_3, which apply to an unstable system. In both cases, in a relatively small time the system automatically selects the optimum value of the coefficients $D_2(z)$, for which the error is a minimum for the given control input $X_0(t)$.

[It is interesting to note that when the load simulator is switched off (X = const.) in the course of a few cycles of the operation of the DCM the quantity $\varepsilon_{av}E$ is reduced to a minimum in the same number as for $X_0(t)$, which indicates the efficiency of the search method used.]

Conclusions

The proposed method of constructing a self-adaptive control programme can easily be realized in a DCM and requires a relatively small number of instructions in the programme.

For the determination of the dynamic properties of the load, in the process of normal operation of the automatic control system with a DCM, it is useful to use a transfer function in the form $W_H(z)$ (the equivalent of a difference equation). Here a good result in the determination of the coefficients of $W_H(z)$ gives a method presented above, which is based on the principle of a 'learning model'.

The experimental investigation showed the efficiency of the self-adaptive control programme, constructed according to the proposed method.

References

[1] TSYPKIN, YA. Z. Theory of pulse systems. *Fizmatgiz* (1958)

[2] TOU, J. *Digital and Sampled-data Control Systems.* 1959. New York

[3] KLUBNIKIN, P. F. Synthesis of control programmes in systems including digital calculating machines. *Automat. Telemech.* Vol. 21, No. 11 (1960)

[4] KALMAN, R. E. Planned self-organizing control systems. *Trans. Amer. Soc. mech. Engrs* No. 57 (1957)

[5] WIDROW, B. Adaptive sampled-data systems. *Automatic and Remote Control.* 1960. London; Butterworths

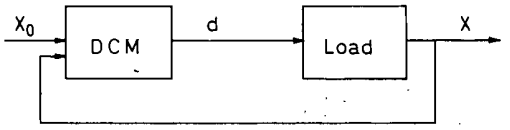


Figure 1. Diagram of the control system

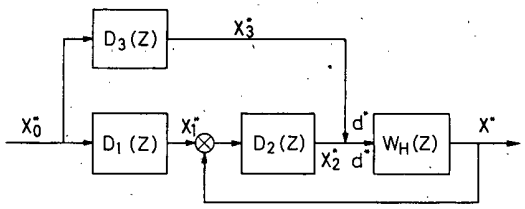


Figure 2. Diagram of the control programme

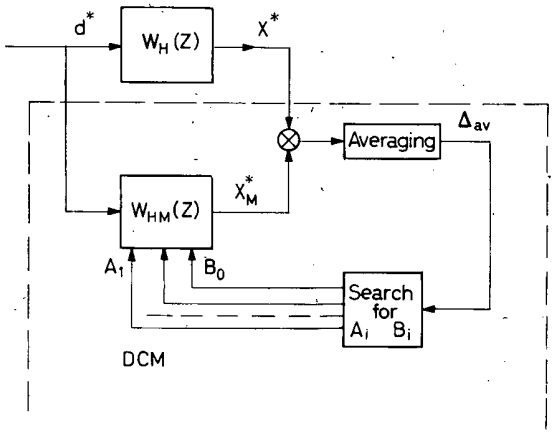


Figure 3: Diagram illustrating the method of determining the coefficients $W_H(z)$, A_i, B_j

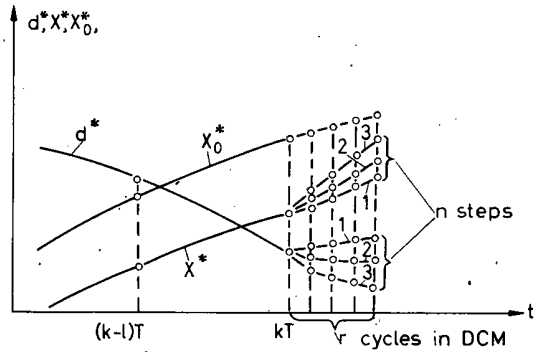


Figure 4. Graphs illustrating the method of constructing a self-adaptive programme

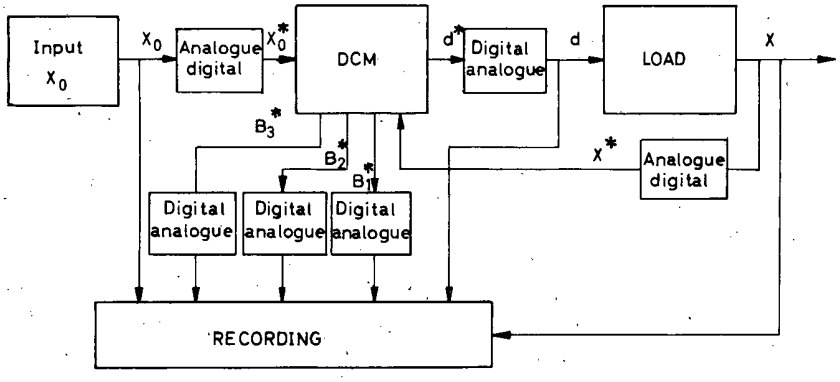


Figure 6. Diagram of control system for performing experiment
Analogue–digital transforms voltage to binary code
Digital–analogue transforms binary code to voltage

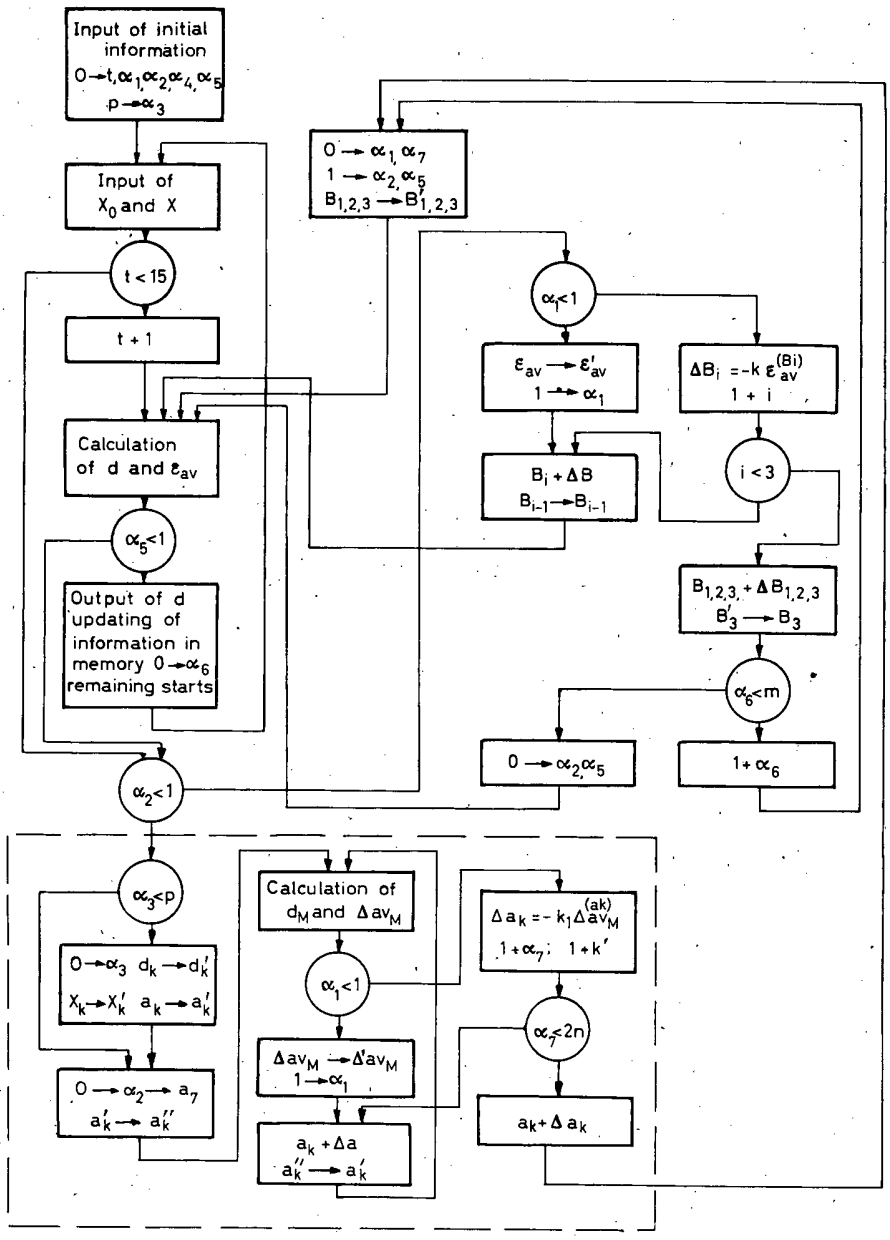


Figure 5. Flow chart of the control programme
p-number ($p > k$) defining input condition of information in X_k, d_k

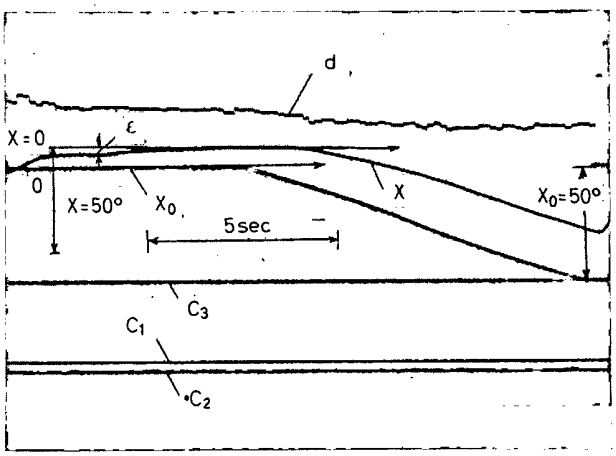


Figure 7. Oscillogram of the evolution of an input control system with constant coefficients $D_2(z)$ with normal control
$C_1 = 0{\cdot}25$, $C_2 = 0{\cdot}5$, $C_3 = -0{\cdot}25$ *selected by numerical means*

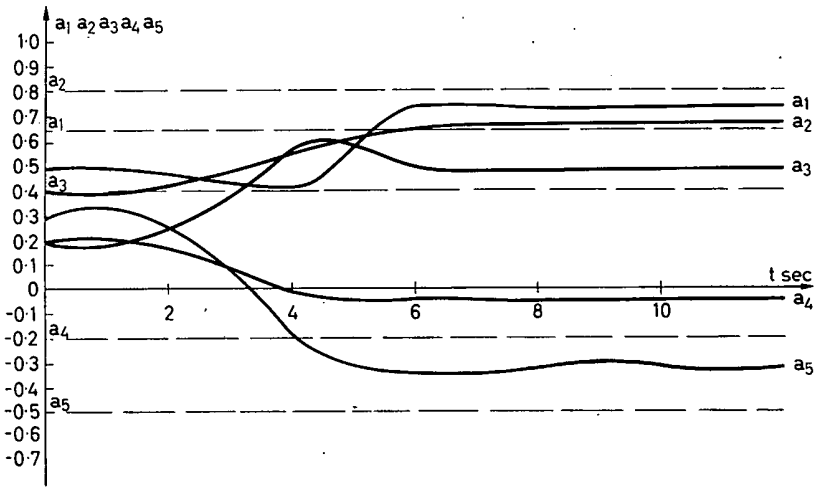


Figure 8. Graph of the change of the coefficients a_i in the process of searching in the determination of $W_{HM}(z)$

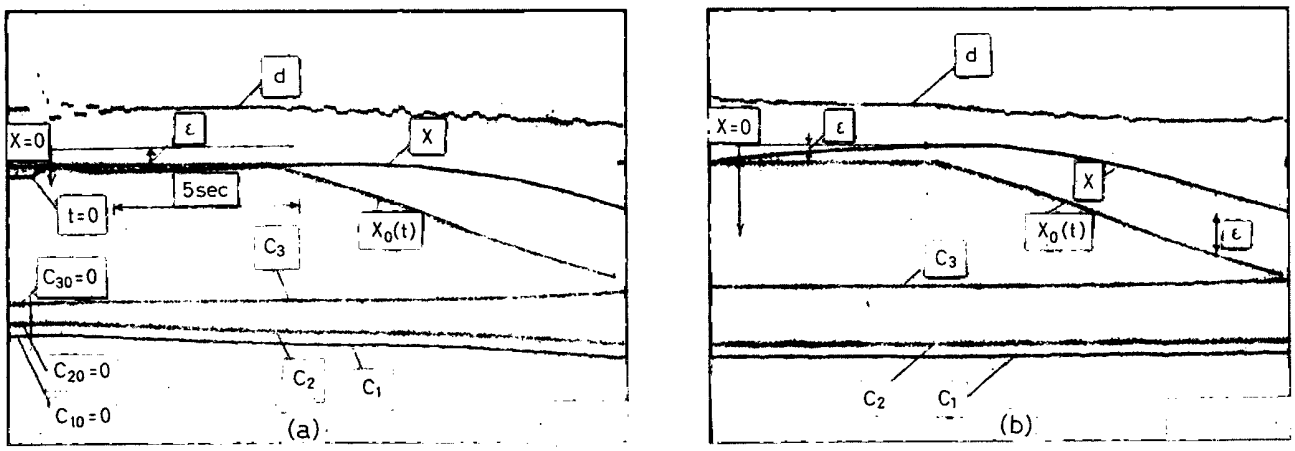


Figure 9. Oscillograms of the operation of the system:
$\Delta C = 2^{-5}$, $K = 0{\cdot}98$, $m = 3$, $r = 2$, $\lambda = 1$ $C_{10} = C_{20} = C_{30} = 0$
(a) portion of initial operation
(b) portion after forming the coefficients

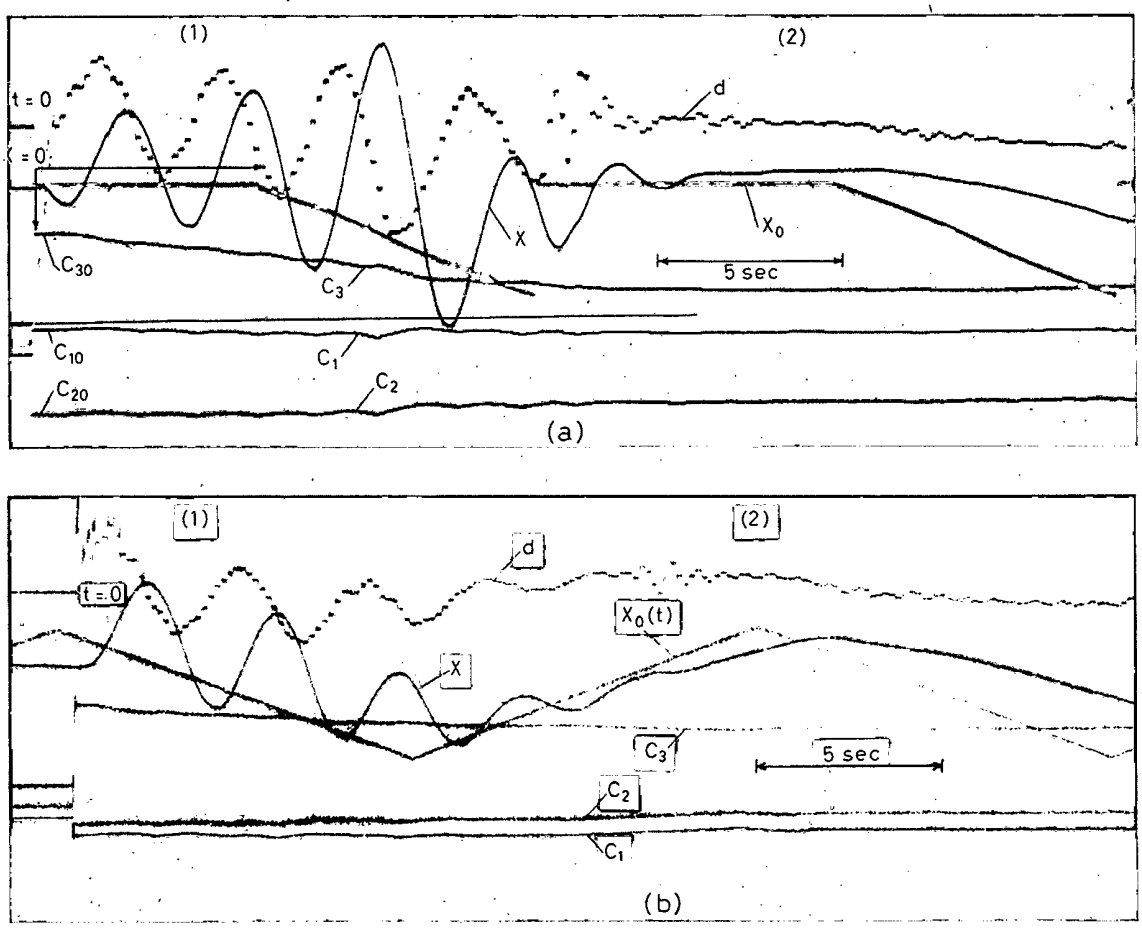


Figure 10. Oscillograms of the operation of the systems: $C = 2^{-5}$, $K = 0{\cdot}98$, $m = 3$, $r = 2$, $\lambda = 1$

(1) *initial operation*

(2) *after forming the coefficients*

(*a*) $C_{10} = -0{\cdot}98 \quad C_{20} = -0{\cdot}25 \quad C_{30} = 0{\cdot}8$

(*b*) $C_{10} = -0{\cdot}9 \quad C_{20} = 0{\cdot}22 \quad C_{30} = 0{\cdot}23$

Non-Linear Programming in the Investigation of Optimal Automatic Control Systems

N. Y. ANDREEV

This paper presents a method of solving a problem in non-linear programming. The essence of the method consists in reducing the set problem to a repeated search for a solution of a linear programming problem and the choice of values for certain additional parameters that are introduced. Non-linear programming problems of a similar nature may be met with in the selection of optimal automatic control systems.

Presently linear programming has deeply penetrated into the techniques used for investigating automatic control systems. Academician Pontryagin's method[1], which determines the optimal control for an automatic system in a number of practically important cases (e.g. the solution of the problem of optimal linear high-speed action), contains a linear programming problem as one of its intermediate stages. Bellman's method of dynamic programming[2], which is of great generality and is also used for investigating automatic control systems, has a linear programming problem as an intermediate stage in a number of cases (when the profit function is linearly dependent on the selected parameters). Linear programming methods are used for solving reliability problems[3], problems of rational tolerances in the production of assemblies[4], and many other problems closely connected with the investigation and development of automatic control systems. It should be noted that in a number of practically important cases the investigation of automatic control systems reduces to a complex problem—a non-linear programming problem, whose solution has so far only been obtained for certain particular cases[4].

This paper puts forward a method of non-linear programming suitable for the solution of a broad range of problems. This method relies essentially on the techniques of linear programming. Therefore the formulation of the linear programming problem is set out below.

As is known[4–6], this problem is expressed in the following manner. It is necessary to find the greatest value of a linear function of n variables $x_1, x_2, \ldots, x_n$

$$L = L(x_1, x_2, \ldots, x_n) = p_1x_1 + p_2x_2 + \ldots + p_nx_n \qquad (1)$$

when the variables are subject to constraints of the form

$$\left.\begin{aligned} &a_{11}x_1 + \ldots + a_{n1}x_n = b_1 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &a_{1m}x_1 + \ldots + a_{nm}x_n = b_m \end{aligned}\right\} \qquad (2)$$

$$\left.\begin{aligned} &d_{11}x_1 + \ldots + d_{n1}x_n \le l_1 \\ &\cdots\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots\cdots \\ &d_{1r}x_1 + \ldots + d_{nr}x_n \le l_r \end{aligned}\right\} \qquad (3)$$

The relations (2) and (3) determine the region G of variation of the variables $x_1, \ldots, x_n$. These conditions can be transformed in such a way that either m or r becomes zero[4]. In actual problems one uses the method of writing the conditions that is most convenient.

In actual problems the function L serves as an index of the quality of the solution. The parameters $x_1, \ldots, x_n$ are characteristic of the object and the investigation, and have various physical significances according to the problem. For example, in solving a problem on high-speed action these parameters appear as control actions.

A geometrical interpretation can be given to the linear programming problem as follows: it is required to find the greatest value of linear function L of the variables $x_1, \ldots, x_n$, whose variation is confined to a region G given in the form of a polyhedron in n-dimensional space.

Efficient techniques have been developed for solving the linear programming problem[4, 5]. But the linear programming method is inapplicable when either the quality index is a non-linear function $F(x_1, \ldots, x_n)$ or the region G of variation of the parameters $x_1, \ldots, x_n$ is determined by non-linear relations between them. Such cases arise, for example, in solving the high-speed action of a system:

1. If the equations of the system include non-linear terms in the control parameters (quality index a non-linear function of the parameters);

2. If the region G of variation of the parameters is determined by non-linear relations, e.g. of the form

$$x_1^2 + \ldots + x_n^2 \le R^2$$

(the region of control forms a hypersphere centred on the origin).

If only the relations defining the region G are non-linear while the quality index is a linear function, one can replace this region by one bounded by relations of the same type as (2) and (3) which coincide accurately enough with the original region (e.g. the hypersphere may be replaced by a polyhedron circumscribed to it). The problem is thus reduced to one of linear programming.

If the quality index is a non-linear function F while the region G is determined by linear relations such as (2) and (3), one can sometimes replace the non-linear function $F(x_1, \ldots, x_n)$ by one that is piecewise linear, and proceed to solve the problem by linear programming methods[4]. But this device cannot always be used, and involves very bulky computation when it is applicable.

In view of what has been said, the following formulation of the non-linear programming problem is of practical and theoretical interest. Let the quality index be a given non-linear function F of the variables $x_1, \ldots, x_n$. Without loss of generality,

it may be considered that the function $F(x_1, \ldots, x_n)$ may be represented as a function Φ of certain linear forms $L_1, L_2, \ldots, L_{k+1}$

$$F(x_1, x_2, \ldots, x_n) = \Phi(L_1, L_2, \ldots, L_{k+1}) \quad (4)$$

where Φ is a given function of the variables $L_1, \ldots, L_{k+1}$;

$$L_i = q_{i0} + q_{i1}x_1 + \ldots + q_{in}x_n \quad (5)$$

the q_{ij} being given numbers for $i = i, 2, \ldots, n$ and $j = 0, 1, 2, \ldots, n$, while $k < n$.

It is required to find the greatest value of the function F under the conditions (2) and (3).

Before proceeding with the solution of this problem, it must be explained why the function F is replaced by Φ. The fact is that in many practical problems the number n of variables is large, and this severely complicates the process of finding a solution. Therefore it is worth while, if at all possible, to go over from the function F of many variables to the function Φ depending on a lesser number of variables L_i. Such a transition, as will be seen from what follows, simplifies the procedure for obtaining a solution.

Two examples are given to illustrate this method of transition to a smaller number of variables.

Example 1—$F(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 + 2\,x_2\,x_3$.

This function of threee variables x_1, x_2 and x_3 can be expressed as a function of two other variables L_1 and L_2:

$$F(x_1, x_2, x_3) = \Phi(L_1, L_2) = \frac{L_1^2 + L_2^2}{2}$$

where $L_1 = x_1 + x_2 + x_3$, $L_2 = x_1 - x_2 - x_3$.
Here $n = 3$ and $k = 1$.

Example 2—$F(x_1, x_2) = x_1^2 + x_2^2$.

This function of two variables cannot be expressed as a function of a lesser number of variables L_i. In this example one may put $L_1 = x_1$ and $L_2 = x_2$. Here $n = 2$ and $k = 1$.

The greatest value of the function F in the region G of variation of the variables $x_1, \ldots, x_n$ as defined by conditions (2) and (3) coincides with the greatest value of the function Φ in the region Q of variation of the variables $L_1, \ldots, L_{k+1}$ as determined also in the final analysis by conditions (2) and (3). The greatest value of Φ may be attained either within the region Q or on its boundary S. Consider each of these cases separately.

First Case

Suppose the function Φ attains a maximum within the region Q. In this event the problem reduces to finding a maximum of a function of $k + 1$ variables. It is known that a necessary condition for Φ to have a maximum is that its partial derivatives should vanish:

$$\frac{\partial \Phi}{\partial L_i} = 0, \quad i = 1, 2, \ldots, k+1 \quad (6)$$

at a certain point in the region Q of the parameter space $(L_1, \ldots, L_{k+1})$.

If Φ is not differentiable everywhere inside Q, some of the conditions (6) may be replaced by these:

$\partial\Phi/\partial L_v$ does not exist, $v = 1, 2, \ldots, m \leq k + 1$.

In the general case the system of eqn (6) may have several solutions. Out of them must be chosen the one that corresponds to the greatest value of Φ. Suppose this solution has been found:

$$L_i = L_{im}, \quad i = 1, 2, \ldots, k+1 \quad (7)$$

Substituting the values (7) of the variables L_i in eqn (5), it is possible to determine the values of the quantities $x_j = x_{jm}$ at which the required greatest value of F is attained. Thus, in this case, the problem is solved by using the normal methods of classical analysis. Conditions (2) and (3) are here used only to reject those maxima of Φ (or F) that do not fall within Q (or G). This first case is rarely met in practice, since the quality index is normally taken as a function F which has no maximum within the region G. The case considered below is of greater practical interest.

Second Case

Suppose the function Φ has no maximum within the region Q, and attains its greatest value on the boundary S of this region. In this case the determination of the greatest value cannot be solved by the techniques of classical analysis, and so the following two-stage method is proposed for solving this problem.

In the first stage one must determine the boundary S of the region Q, while in the second, one finds the greatest value of the function Φ on S. Here one may make use of the ideas and techniques developed by the author[7, 8], applying them to a problem of a different nature.

To determine the boundary S one may proceed in the following manner. For fixed values of the variables

$$L_1 = C_1, L_2 = C_2, \ldots, L_k = C_k \quad (8)$$

one must find the greatest (and least) value of L_{k+1} (see *Figure 1*, where $k = 1$).

Since the greatest and least values of L_{k+1} are determined by similar means, from now on only the greatest values of L_{k+1} are mentioned (i.e. only one half-branch of S is dealt with).

It follows that, to find one point on S, one must obtain the greatest value of the linear form L_{k+1} under conditions (2), (3) and (8). This is a typical linear programming problem. Conditions (8) have essentially changed nothing in conditions (2) and (3); the number of equations has merely increased by k. Taking various values of the parameters $C_1, C_2, \ldots, C_k$, one can also derive the points on S corresponding to them. If these points are chosen so as to cover the whole of S densely enough, the first stage of the problem may be considered solved.

Now it is necessary to solve the second stage of the problem, i.e. to find the greatest value of Φ on S. This is easily solved if the number k of dimensions of S is small. In this case the greatest value of Φ can be determined approximately by comparing the values of Φ at the nodes of a network formed by discrete values of the numbers $C_1, C_2, \ldots, C_k$. If the number k of dimensions of S is large, producing a close network of values of Φ on S becomes an extremely laborious task, which cannot always be carried out in a reasonable time even by the use of modern high-speed computer techniques.

In this case the determination of the greatest value of Φ reduces to finding the maximum of the function

$$f = f(C_1, C_2, \ldots, C_k) = \Phi[C_1, \ldots, C_k, L_{k+1}(C_1, \ldots, C_k)] \quad (9)$$

where $L_{k+1}(C_1, \ldots, C_k)$ is the greatest (least) value of the linear form L_{k+1} under conditions (2), (3) and (8). The greatest value of f on S in general coincides with the maximum of this function, i.e. isattained within the region of variation of the parameters C_i.

To determine the maximum of $f = f(C_1, \ldots, C_k)$ use can be made of the method of most rapid descent[8, 9]. The combination of the method given above (which leads to the boundary S of the region Q, and to the function f) and the method of most rapid descent (leading to the maximum of f) makes it possible to avoid the computation of values of f at a large number of points densely covering the whole region S of variation of the variables $C_1, \ldots, C_k$, and to replace these bulky calculations by more economic ones according to the following plan.

Let a first approximation to the variables

$$C_1 = C_{11}, C_2 = C_{21}, \ldots, C_k = C_{k1}$$

be chosen from any considerations. To this corresponds a value of the function $f_1 = f(C_{11}, C_{21}, \ldots, C_{k1})$. Now the direction of the gradient of f at this point is determined, which as is known is given by a vector in the space $G = (C_1, \ldots, C_k)$, whose projections on the $C_1, C_2, \ldots, Ck$ axes are respectively

$$\frac{\partial f}{\partial C_1}, \frac{\partial f}{\partial C_2} \cdots \frac{\partial f}{\partial C_k}$$

The partial derivatives $\partial f/\partial C_i$ may be derived analytically if one has succeeded in obtaining a simple analytical expression for f.

But one cannot count on this, since normally the expression for f is complicated and, what is more, cannot be derived in explicit form. Thus in the general case the derivatives $\partial f/\partial C_i$ must be obtained approximately as the ratio of finite differences

$$\frac{\partial f}{\partial C_i} \simeq \frac{\Delta f_i}{\Delta C_i}$$

where $\Delta f_i = f(C_{11}, \ldots, C_{(i-1)1}, C_{i1} + \Delta C_i, C_{(i+1)1}, \ldots, C_{k1}) - f(C_{11}, \ldots, C_{k1})$.

After determining the gradient of f at the point $(C_{11}, \ldots, C_{k1})$, a displacement in the space G is made along this gradient vector, i.e. the values of f are considered for the following values of the variables:

$$C_1 = C_{11} + \frac{\partial f}{\partial C_1} \cdot \varepsilon, \; C_2 = C_{21} + \frac{\partial f}{\partial C_2} \cdot \varepsilon, \; \ldots, \; C_k = C_{k1} + \frac{\partial f}{\partial C_k} \cdot \varepsilon$$

where the $\partial f/\partial C_i \, (i = 1, 2, \ldots, k)$ are evaluated at $C_1 = C_{11}$, $C_2 = C_{21}, \ldots, C_k = C_{k1}$.

The displacement in the chosen direction is terminated at the value $\varepsilon = \varepsilon_1$ at which the function

$$\xi_1(\varepsilon) = f\left(C_{11} + \frac{\partial f}{\partial C_1}\varepsilon, \ldots, C_{k1} + \frac{\partial f}{\partial C_k}\varepsilon\right)$$

reaches a maximum. This maximum of $\xi_1(\varepsilon)$ may be determined graphically (see *Figure 2*).

The values of the variables

$$C_1 = C_{12} = C_{11} + \frac{\partial f}{\partial C_1}\varepsilon_1, \ldots, C_k = C_{k2} = C_{k1} + \frac{\partial f}{\partial C_k}\varepsilon_1$$

are taken as the second approximation. The value of the function $f = f_2 = f(C_{12}, \ldots, C_{k2})$ is taken as the second approximation to f.

Then the third and succeeding approximations to the variables $C_1, \ldots, C_k$ and the function f are obtained by the method given above.

The $(v + 1)$th approximation is given by the formulae:

$$C_{1(v+1)} = C_{1v} + \frac{\partial f}{\partial C_1} \cdot \varepsilon_v, \ldots, C_{k(v+1)} = C_{kv} + \frac{\partial f}{\partial C_k} \cdot \varepsilon_v$$

where $\varepsilon = \varepsilon_v$ corresponds to a maximum of the function

$$\xi^v_{(\varepsilon)} = f\left(C_{1v} + \frac{\partial f}{\partial C_1}\varepsilon, \ldots, C_{kv} + \frac{\partial f}{\partial C_k}\varepsilon\right)$$

$$f_{v+1} = f(C_{1(v+1)}, C_{2(v+1)}, \ldots, C_{k(v+1)})$$

The process of finding the maximum of f is terminated when two successive approximations to f differ by a negligible amount.

In the general case the function f may have several maxima, and it is necessary to find the greatest of these. It should be noted that in the general case the greatest value of f is attained within the region S of variation of the parameters $C_1, C_2, \ldots, C_k$, i.e. it coincides with a maximum of this function. Only in rare individual cases is the greatest value of f attained on the boundary of the region S. This assertion follows from the fact that in the general case the function $F(x_1, x_2, \ldots, x\hat{n})$ reaches its greatest value on a face, and not at a vertex, of the polyhedron defined by conditions (2) and (3).

Based on the above, the following sequence of operations can now be recommended for determining a maximum of the function f.

(*a*) Choose the first approximation to the variables $C_1 = C_{11}, \ldots, C_k = C_{k1}$.

(*b*) Compute the value of the first approximation to the function $f = f_1 = f(C_{11}, \ldots, C_{k1})$.

(*c*) Evaluate the components of the gradient vector of f at the first approximation point:

$$\frac{\partial f}{\partial C_1}, \ldots, \frac{\partial f}{\partial C_k}$$

(*d*) Calculate the function

$$\xi_1(\varepsilon) = f\left(C_{11} + \frac{\partial f}{\partial C_1}\varepsilon, \ldots, C_{k1} + \frac{\partial f}{\partial C_k}\varepsilon\right)$$

for increasing values of the parameter $\varepsilon = \Delta\varepsilon \cdot l$, where $l = 1, 2, \ldots$. The increment $\Delta\varepsilon$ is chosen in accordance with the peculiarities of f that become evident during the process of computation: the more gentle the variation in f, the larger can $\Delta\varepsilon$ be taken.

(*e*) Determine the value of the parameter $\varepsilon = \varepsilon_1$ that makes the function $\xi_1(\varepsilon)$ a maximum.

(*f*) Determine the second approximation

$$C_{12} = C_{11} + \frac{\partial f}{\partial C_1}\varepsilon_1, \ldots, C_{k2} = C_{k1} + \frac{\partial f}{\partial C_k}\varepsilon_1$$

(*g*) Evaluate the second approximation to f:

$$f_2 = f(C_{12}, C_{22}, \ldots, C_{k2}).$$

(*h*) Calculate the difference between the two successive approximations to f, i.e. $f_2 - f_1$.

This sequence is continued until the difference

$$f_{t+1} - f_t = f(C_{1(t+1)}, \ldots, C_{k(t+1)}) - f(C_{1t}, \ldots, C_{kt})$$

becomes negligibly small. Ordinarily the number of approximations that have to be taken when using this technique is not great. The computations involved can readily be programmed for a computer dealing with finite differences.

One may naturally wonder whether the method of most rapid descent cannot be applied directly to determining the greatest value of the function $F(x_1, \ldots, x_n)$ under conditions (2) and (3). In principle, this approach is also possible, but it leads to substantially more complex calculations in the cases where (*a*) the number n of variables $x_1, \ldots, x_n$ is significantly greater than the number k of variables $C_1, \ldots, C_k$, and (*b*) the number of inequalities (and equations) in conditions (2) and (3) is large.

The considerable increase in the volume of computation in the first case needs no explanation. In the second case, it arises from the fact that the direct application of the method of most rapid descent here requires that at each step of the calculation, when ε is increased by $\Delta\varepsilon$, one has also to check whether or not conditions (2) and (3) are satisfied. Also the transition from one face to another of the polyhedron defined by (2) and (3) involves a change in the form of a function of $n - r$ variables.

This complicates the programming of the computation. It follows that the volume of work in deriving each approximation increases, and so does the number of approximations.

When the number of inequalities in (3) is small and $k + 1 = n$, both the methods become roughly equal in time-consumption.

These two different cases have been considered above: (*a*) The greatest value of F (and Φ) is attained within the region of variation of the variables $x_1, \ldots, x_n$ (or $L_1, \ldots, L_{k+1}$), and (*b*) the greatest value of F (and Φ) is attained on the boundary of this region, the function having no maximum within the region G (or Q).

The case may arise [al though also improbable, as (*a*) above] where the function F (or Φ) has a maximum within the region G (or Q), but attains its greatest value on the boundary of this region. Consequently in this case the maximum of Φ has to be found and compared with the greatest value of this function reached on the boundary S, and the greater of the two has to be chosen.

It may be expected that the techniques of solving non-linear programming problems will develop in the future, and that experience in this field will accumulate. Therefore it is worth making the following more general statement of the problem.

Let there be a method for determining the greatest (and least) value of the function $\Psi(x_1, \ldots, x_n)$ under conditions (2) and (3) that are imposed on the region of variation of the variables $x_1, \ldots, x_n$. It is necessary to find the greatest value of the function

$$F(x_1, \ldots, x_n) = \Phi(\Psi, L\hat{1}, \ldots, L_k) \qquad (10)$$

where $L_p = q_{p0} + q_{p1} x_1 + \ldots + q_{pn} x_n, p = 1, 2, \ldots, k, k < n$, and conditions (2) and (3) are satisfied.

Consider Φ as a function of the $k + 1$ parameters $\Psi, L_1, \ldots, L_k$. The greatest value of this function may be attained either within the region Q of variation of these variables or on its boundary.

If the greatest value of Φ is reached within Q (an improbable case in practice), then the problem reduces to finding the maxima of this function, which are determined by the equations:

$$\frac{\partial\Phi}{\partial\Psi} = 0 \qquad \frac{\partial\Phi}{\partial L_p} = 0 \qquad p = 1, 2, \ldots, k \qquad (11)$$

These equations enable one to determine the values of the functions $\Psi = \Psi_0, L_1 = L_{10}, \ldots, L_k = L_{k0}$, which correspond to a maximum of Φ. If the solution of eqn (11) is not unique, then one must choose from all its solutions the one that corresponds to the greatest of the maxima of Φ. From the relations

$$\Psi(x_1, \ldots, x_n) = \Psi_0$$

$$q_{p0} + q_{p1} x_1 + \ldots + q_{pn} x_n = L_{p0}, \quad p = 1, 2, \ldots, k \qquad (12)$$

one determines the values of the variables $x_{10}, x_{20}, \ldots, x_{n0}$ corresponding to the greatest value of the function Φ. In the general case the solution of the system (12) is not unique.

If, however, the greatest value of Φ is reached on the boundary S of the region Q (which is more likely in practical cases), then it is desirable to solve the problem as stated in two stages. First one must find the boundary S, and then determine the greatest value of Φ on it. In determining the boundary S, it is necessary to take given values of the linear forms

$$L_1 = C_1, L_2 = C_2, \ldots, L_k = C_k \qquad (13)$$

and then determine the greatest and least values of the function $\Psi(x_1, x_2, \ldots, x_n)$ under conditions (2), (3) and (13).

It has been pointed out above that there is a method for solving this problem [the addition of (13) does not in principle alter conditions (2) and (3)]. Taking various given values of the parameters $C_1, \ldots, C_k$ one may obtain the corresponding values:

$$\Psi_1 = \Psi(C_1, C_2, \ldots, C_k)$$

$$\Phi = \Phi[\Psi(C_1, \ldots, C_k), C_1, \ldots, C_k] = f(C_1, \ldots, C_k)$$

$$x_i = x_i(C_1, C_2, \ldots, C_k) \quad i = 1, 2, \ldots, n$$

The second stage of the solution consists in the determination of the maximum of $f = f(C, \ldots, C_k)$ and the values of the variables

$$x_1 = x_{10}, x_2 = x_{20}, \ldots, x_n = x_{n0}$$

corresponding to this maximum. This part of the solution is carried out in exactly the same way as for the first statement of the problem. A simple example is now given to explain the technique that has been proposed for solving the non-linear programming problem.

Example

Determination of the greatest value of the function

$$F(x_1, x_2, x_3) = x_1(x_2 + x_3)$$

under the conditions:

$$x_1 + 2x_2 + 3x_3 \leq 60$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0$$

The given function F of the three variables x_1, x_2 and x_3 can be expressed as a function Φ of two linear forms L_1 and L_2:

$$L_1 = x_2 + x_3, L_2 = x_1$$

with

$$\Phi(L_1, L_2) = L_1 \cdot L_2$$

In this example Φ is a monotone increasing function. Hence it has no maximum, and attains its greatest value on the boundary S. Following the procedure set out above, one determines the boundary S of the region Q of variation of the linear forms L_1 and L_2. In this case the boundary is a certain curve (a one-dimensional domain).

In order to find S, it is necessary to take various given values of the linear form L_1 and to evaluate for each the greatest and least values of L_2. The question arises of how to choose these given values of L_1. This question is easily answered. The greatest and least values of L_1 under the above conditions are readily obtained by linear programming methods and are:

$$0 \leq L_1 \leq 30$$

Taking a certain value $L_1 = C_1$, where $0 \leq C_1 \leq 30$, the greatest value of L_2 is found (the least value of L_2 is of no interest in this example, since Φ is a monotone increasing function in the variable L_2). This greatest value is easily obtained by linear programming methods (or by other means), and may be expressed in terms of C_1 in the following form:

$$L_2 = 60 - 2C_1$$

The function Φ can be expressed in terms of the parameter C_1 as follows over the section of S that is being considered:

$$\Phi = C_1(60 - 2C_1)$$

It can readily be seen that the function Φ on the boundary S attains a maximum at $C_1 = C_{10} = 15$.

Now it is easy to determine all the quantities of interest:

$$\Phi_0 = F_0 = 15(60 - 2 \cdot 15) = 450$$

$$x_{10} = 60 - 2 \cdot 15 = 30$$

The values of x_{20} and x_{30} are obtained from the equations:

$$x_2 + x_3 = 15$$

$$30 + 2x_2 + 3x_3 = 60 \quad \text{(see above conditions)}$$

Solution of these equations gives the following values for x_{20} and x_{30}:

$$x_{20} = 15, \quad x_{30} = 0$$

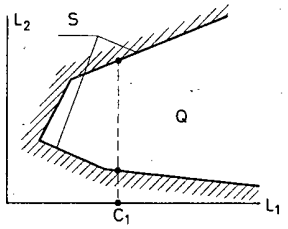


Figure 1

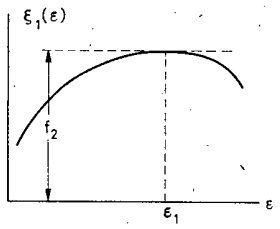


Figure 2

References

[1] PONTRYAGIN, L. S. [redacted] *The Mathematical Theory of Optimal Process.* 1961. Moscow; Fizmatgiz

[2] BELLMAN, R. Dynamic programming. (Translated into Russian.) *Inostr. Lit.* 1960

[3] SANDLER, D. Paper in the collected reports of the 5th U.S.A. Symposium, 1959

[4] YUDIN, D. V., and GOL'SHTEYN, YE. G. *Problems and Methods of Linear Programming.* 1961. Sovetskoye Radio

[5] GASS, S. *Linear Programming.* 1961. Moscow; Fizmatgiz

[6] KRASOVSKIY, A. A., and POSPELOV, G. S. *Fundamentals of Automation and Technical Cybernetics.* 1962

[7] ANDREEV, N. Y. A method of determining the optimum dynamic system from the criterion of extreme of a functional which is a given function of several other functionals. *Automatic and Remote Control.* 1961. London; Butterworths, Vol. 2, p. 707

[8] ANDREEV, N. Y. Determination of an optimal dynamic system from the criterion of a functional of partial form. *Automat. Telemech.; Moscow* 18, No. 7 (1957)

[9] KANTOROVICH, L. V. Functional analysis and applied mathematics. *Usp. Mat. Nauk* 3, No. 6 (1948)

The Approximate Calculation of a Class of Automatic Systems with Forced Parameter Optimization

YU. I. ALIMOV

Introduction

§ 1. This paper considers an automatic system (hereafter called System A) that consists of linear continuous filters Φ_1 and Φ_2 connected in parallel, with a test signal $\theta(t)$ at the input and closed-loop astatic systems for adjusting the parameters $X = (X_1, \ldots, X_N)$ of filter Φ_2 (see *Figure 1*). The self-adjusting circuit includes a detector $\mathscr{D}$ of the error signal $\varepsilon(t)$, phase discriminators $\Phi\mathscr{D}_i$, averaging filters $W_{\phi i}$ and integrating networks. The control actions in the parameter-adjusting circuits are formed by using a search modulation $\mu\Delta x(t)$ of the parameters. The defined parameters Y_1 and Y_2 of filters Φ_1 and Φ_2 respectively vary with time according to a law that is only known approximately beforehand.

In practice, the following variants of System A are most often met.

(1) $Y_1 = \text{const.}$, $X_2 = Y_2(t)$. The filter Φ_1 is a stationary calibration display unit, while the filter Φ_2 is an automatic system with extremal adjustment of its correcting elements, compensating given to a extent the drift of the parameters $Y_2(t)$[1–4] or the variation in the form of the external action $\theta(t)$[5].

(2) $Y_1 = Y_1(t)$, $Y_2 = \text{const}$. Filter Φ_1 is a controlled plant with variable dynamic properties, while filter Φ_2 is a learing model of this plant[6].

Of course, the general case $Y_1 = Y_1(t)$, $Y_2 = Y_2(t)$ is also possible in practice; for example, a calibration display unit Φ_1 with programmed parameter variation.

§ 2. In Part I of the paper the small-parameter method is used in deriving enough general approximate equations for the processes of self-adjustment in System A under the assumption that the amplitudes of the search signals $\mu\Delta x(t)$ are small. The equations take account of the limited memory of filters ϕ_1 and ϕ_2, and cover the case of any given explicit test and search actions. The control signals in the self-adjusting circuits are expressed in terms of the frequency characteristics of filters ϕ_1 and ϕ_2 and the spectra of signals $\theta(t)$ and $\mu\Delta x(t)$. In Part II the general equations of motion for System A are simplified, taking the assumption that the search signals $\mu\Delta x(t)$ are sinusoidal. Then, as a simplified mathematical abstraction, the case of an almost periodic action $\theta(t)$ is examined in detail in Part III. A very simple analysis of the relevant equations of motion shows the desirability, with a high-frequency sinusoidal signal $\mu\Delta x_i(t)$, of using, in the phase discriminator, a reference voltage phase chifted with respect to this signal, which permits one to make use of the extra useful information carried by the quadrature component of the search-frequency signal, by analogy with the practice, in radio engineering, of using amplitude and phase modulation simultaneously[7]. Part IV uses the example of awhite-noise test signal to show that the equations derived may also be applied to the description of System A with stochastically defined signals $\theta(t)$, without relying on the hypothesis of the closeness of random processes in the system to stationary ergodic ones. There is a brief discussion of the relation between the results derived here and those in previous papers[1–6]. Some attention is also devoted to quasi-stationary modes of self-adjusting operation.

In conclusion it should be stressed that all the design examples quoted have been chosen to be simple as far as possible, and that the main emphasis is on the physical interpretation and qualitative analysis of the mathematical relations derived.

I. Derivation of General Equations of Motion for the Self-adjusting System Considered

§ 3. The most important of the assumptions, under which the equations for the processes of self-adjustment in System A are derived below, are first set out:

(*a*) The amplitudes of the search-modulation signals are considered small, and to emphasize this they are denoted by $\mu\Delta x(t)$ where μ is a small parameter.

(*b*) It is assumed that System A starts to operate at a certain instant $t = t_0$, having been in an equilibrium condition up to that time, and thus the output quantity of filter Φ_i $(i = 1, 2)$ is determined by the relation

$$\xi_i(t) = \int_{t_0}^{t} \theta(\tau) K_i(t, \tau)\, d\tau, \quad i = 1, 2 \tag{1}$$

where $K_i(t, \tau)$ is the weighting function of filter Φ_i.

Equation (1) is expressed in the form

$$\xi_i(t) = \int_{t_0}^{t-T} \theta(\tau) K_i(t, \tau)\, d\tau + \int_{t-T}^{t} \theta(\tau) K_i(t, \tau)\, d\tau, \quad i = 1, 2$$
$$t_0 < t - T < t \tag{2}$$

Let the filters Φ_i be stable. Then if

$$|\theta(t)| < \text{const} \quad (-\infty < t < +\infty) \tag{3}$$

it may be considered that for a certain sufficiently large T the first integral in eqn (2) is negligibly small, and

$$\xi_i(t) \simeq \int_{t-T}^{t} \theta(\tau) K_i(t, \tau)\, d\tau, \quad i = 1, 2 \tag{4}$$

In other words, this means that by the instant t information on the state of filters Φ_i and on the values of the signal $\theta(\tau)$ at

instants $\tau > t - T$ is practically completely lost, and the value of $\xi_i(t)$ may be identified with the reaction of filter Φ_i to the signal

$$\theta_T(\tau)=\begin{cases}\theta(\tau) \text{ for } t-T<\tau\leq t\\ 0 \text{ outside that interval}\end{cases} \quad (5)$$

assuming that $\xi_i(\tau) \equiv 0$ for $\tau < t - T$.

The conditional nature of any choice of a numerical value for T matches the complexity of the actual situation: the 'memory' T of a linear system depends substantially on the criterion chosen and on many factors that are often not subject to any sort of accurate quantitative calculation (on the structure of the signal $\theta(t)$ within the bounds of the natural and easily enough controlled restriction (3), on the level of fluctuating disturbance in the system, etc.). If filter Φ_i is near to the stability boundary in parameter space, then of course $T \to \infty$. If stability is lost then (4) is not true even for $T = \infty$, and strictly speaking, one cannot apply either the theory developed below, which takes no account of the initial perturbations always existing in a system, or the theories of Krasovskiy[2], Kazakov[3] and Varygin[4].

(*c*) It is also considered that the variation in parameters $Y_1(t)$, $Y_2(t)$ and $X(t)$ over the time interval T may be neglected. The time-dependence of the frequency characteristics $W_1(j\omega) = W_1(j\omega, t)$ and $W_2(j\omega) = W_2(j\omega, t)$ of filters Φ_1 and Φ_2 (with $\Delta x(t) \equiv 0$) is only expressed in the taking of the values of parameters $Y_1(t)$, $Y_2(t)$ and $X(t)$ as 'frozen' at the given instant t:

$$Y_1(\tau)\simeq Y_1(t),\ Y_2(\tau)\simeq Y_2(t),\ X(\tau)\simeq X(t) \text{ for } t-T<\tau<T \quad (6)$$

(*d*) Finally, for the sake of definition, it is assumed that the state of filter ϕ_2 is described by the ordinary differential equation

$$\sum^{n} a_{2,k}(Y_2, X+\mu\Delta x)\, D^k \xi_i(t)=\sum^{m} b_{2,l}(Y_2, X+\mu\Delta x)\, D^l \theta(t) \quad (7)$$

$$D\equiv\frac{\mathrm{d}}{\mathrm{d}t},\quad m\leq n$$

with coefficients $a_{2,k}(Y_2, X)$ and $b_{2,k}(Y_2, X)$ that are analytic in X. It is evident that

$$W_2(j\omega)=R_2(j\omega)\cdot Q_2^{-1}(j\omega) \quad (8)$$

where

$$Q_2(D)=\sum_{k=1}^{n} a_{2,k}(Y_2, X)\, D^k$$

$$R_2(D)=\sum_{l=1}^{m} b_{2,l}(Y_2, X)\, D^l \quad (9)$$

Given assumptions (*a*), (*b*) and (*c*) the proposed method of calculation can be generalized without much complication to the case where pure delays are present in the filter Φ_2 under adjustment.

§ 4. It is observed that assumption (*b*) allows one to make calculation in a frequency region bounded only by consideration of the 'shortened' present spectrum[8]

$$\theta_T(j\omega, t)=\int_{t-T}^{t}\theta(\tau)\,\mathrm{e}^{-j\omega\tau}\,\mathrm{d}\tau=\int_{-\infty}^{t}\theta_T(\tau)\,\mathrm{e}^{-j\omega\tau}\,\mathrm{d}\tau \quad (10)$$

of the signal $\theta(t)$. Thus, in particular, taking into account the quasi-stationary nature of the filter $W_1(j\omega, t)$ the following relations are obtained for $\xi_1(t)$:

$$\xi_1(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\xi_1(j\omega, t)\cdot\mathrm{e}^{j\omega t}\,\mathrm{d}\omega \quad (11)$$

where

$$\xi_1(j\omega, t)\simeq W_1(j\omega, t)\cdot\theta_T(j\omega, t) \quad (12)$$

Considering, instead of normal spectra, the 'shortened' present spectra of the type in (10) and (12), generally one can reflect more accurately in a mathematical model the actual situations that arise in the experimental development of System A, and also simplify mathematical operations on the spectra of the signals $\theta(t)$ and $\xi_i(t)$ in those cases where the Fourier integrals for these functions over the interval $(-\infty, t)$ diverge. This approach turns out, in particular, to be very convenient for the examination of non-ergodic random processes in System A, as it gives a natural transition to the description of the system in terms of spectral power densities (see Part IV).

Since this paper only considers explicit (and, what is more, only harmonic) search signals $\mu\Delta x(t)$, from now on in order to simplify the text the 'full' spectrum is used as a convenient, if less accurate, mathematical abstraction

$$\mu\Delta x_i(j\omega)=\mu\int_{-\infty}^{\infty}\Delta x_i(\tau)\cdot\mathrm{e}^{-j\omega\tau}\,\mathrm{d}\tau \quad (13)$$

of the search signal.

§ 5. A solution $\xi_2(t)$ to eqn (7) is looked for in the form of a series

$$\xi_2(t)=\xi_{20}(t)+\mu.\xi_{21}(t)+\ldots \quad (14)$$

all the analysis below being taken only with the accuracy of magnitudes of the first order of smallness with respect to the quantity μ [obviously one way of making the theory more accurate is to take account of more terms in (14)]. Using the normal procedure[9] for the small-parameter method, the following equation for sequential calculation of the quantities $\xi_{20}(t_0(t)$ and $\xi_{21}(t)$ are obtained from (7)–(9):

$$Q_2(D)\,\xi_{20}(t)=R_2(D)\,\theta(t),\ D\equiv\frac{\mathrm{d}}{\mathrm{d}t} \quad (15)$$

$$Q_2(D)\,\xi_{21}(t)=\sum_{i=1}^{N}\Delta x_i(t)\left[\frac{\partial R_2(D)}{\partial x_i}\theta(t)-\frac{\partial Q_2(D)}{\partial x_i}\xi_{20}(t)\right] \quad (16)$$

It is easily seen that given assumption (*b*) the memory of the linear system (16) should be considered as limited to the time interval T. Hence, taking into account the quasi-stationary nature of the filters $W_1(j\omega, t)$ and $W_2(j\omega, t)$ and the identity $\partial W_1/\partial X_i \equiv 0$, the following expression is found for the 'shortened' present spectrum of the error

$$\varepsilon(t) = \xi_{20}(t) - \xi_1(t) - \mu\xi_{21}(t) \quad (17)$$

$$\varepsilon(j\omega, t)\simeq W(j\omega)\,\theta_T(j\omega, t)$$
$$+\frac{\mu}{2\pi}Q_2^{-1}(j\omega)\sum_{i=1}^{N}\int_{-\infty}^{\infty}\frac{\partial W(jv)}{\partial x_i}Q_2(jv)\,\theta_T(jv, t)\,\Delta x_i(j(\omega\cdot v))\cdot\mathrm{d}v \quad (18)$$

where $W(j\omega = W_2(j\omega) - W_1(j\omega)$

while the spectra $\theta_T(j\omega, t)$ and $\Delta x_i(j\omega)$ are defined by eqns (10) and (13).

Furthermore, in accordance with the circuit shown in *Figure 1*, one obtains for the X_i-adjusting network

$$DX_i = W_{\varphi i}(D)\,\varphi\{\varepsilon(t)\}\,\Delta x_i(t), \quad D \equiv \frac{\mathrm{d}}{\mathrm{d}t} \tag{19}$$

where $\varphi(\varepsilon)$ is the detector characteristic, while

$$\varepsilon(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varepsilon(j\omega, t)\cdot \mathrm{e}^{j\omega t}\,\mathrm{d}\omega \tag{20}$$

The approximate system of eqns (17)–(20) that has been obtained describes a very wide class of self-adjusting operating conditions for System A.

The following points are stressed:

(1) These equations, written in terms of the frequency characteristics, are differential equations (generally speaking, non-linear) and in a number of cases are capable of more effective investigation than the integro-differential equations derived by Krasovskiy[2] and Varygin[4] in terms of weighting functions.

(2) In distinction to the previous papers quoted[2–4], the derivation of eqns (17)–(20) does not rely on the assumption that the weighting function, and consequently also the transfer function, of filter Φ_2 is actually a function rather than a functional of the signals $\mu\Delta x(t)$.

II. Simplification of the General Self-adjustment Equations for the Case of a Harmonic Search Modulation and a Square-law Detector

§ 6. If during the whole time of operation of System A the search $\mu\Delta x(t)$ are nearly harmonic, then it is convenient to consider that

$$\Delta x_i(t) = \Delta_i \cos \Omega_i t \text{ for } -\infty < t < \infty, \quad i = 1, \ldots, N \quad (\Omega_i < \Omega_{i+1}) \tag{21}$$

Then in accordance with (13)

$$\Delta x_i(j\omega) = \pi\left[\delta(\omega+\Omega_i) + \delta(\omega-\Omega_i)\right] \tag{22}$$

Substituting (22) in (17) and using the known properties of δ functions, one finds:

$$\begin{aligned}\varepsilon(j\omega, t) &= W(j\omega)\,\theta_T(j\omega, t)\\ &\frac{1}{2}\mu\, Q_2^{-1}(j\omega)\sum_{i=1}^{N}\\ &\Delta_i\left[\frac{\partial W\{j(\omega+\Omega_i)\}}{\partial x_i} Q_2\{j(\omega+\Omega_i)\}\,\theta_T\{j(\omega+\Omega_i), t\}\right.\\ &\left.+\frac{\partial W\{j(\omega-\Omega_i)\}}{\partial x_i} Q_2\{j(\omega-\Omega_i)\}\,\theta_T\{j(\omega-\Omega_i), t\}\right]\end{aligned} \tag{23}$$

Then let

$$\begin{aligned}W(j\omega) &= |W(j\omega)|\,\mathrm{e}^{j\varphi}\,(\varphi = \varphi(\omega)),\\ \theta_T(j\omega, t) &= |\theta_T(j\omega, t)|\,\mathrm{e}^{j\alpha} \qquad (\alpha = \alpha(\omega))\end{aligned} \tag{24}$$

Taking into account the even nature of the amplitude spectrum and the odd nature of the phase spectrum in (24), one can readily deduce from (20) and (22) the following expression for the error $\varepsilon(t)$:

$$\begin{aligned}\varepsilon(t) \simeq \frac{1}{\pi}\int_0^{\infty}\Big\{&|W(j\omega)|\cos(\omega t+\varphi+\alpha)\\ &+\frac{1}{2}\mu\sum_{i=1}^{N}\Delta_i\left[ReC_i\cos(\omega t+\alpha)\right.\\ &\left.-\mathrm{Im}C_i\sin(\omega t+\alpha)\right]|\theta_T(j\omega, t)|\Big\}\,\mathrm{d}\omega\end{aligned} \tag{25}$$

$$C_i = \frac{\partial W(j\omega)}{\partial x_i} Q_2(j\omega)\left[\frac{\mathrm{e}^{-j\Omega_i t}}{Q_2\{j(\omega-\Omega_i)\}} + \frac{\mathrm{e}^{-j\Omega_i t}}{Q_2\{j(\omega+\Omega_i)\}}\right] \tag{26}$$

In calculating the passage of the signal $\varepsilon(t)$ through the detector $\mathscr{D}$, it is convenient first to separate, in each term of the integrand in (24) that is enclosed, in square brackets the components in phase with the signal $\cos(\omega t + \varphi + \alpha)$ and those in quadrature with it. One obtains as a result:

$$\begin{aligned}\varepsilon(t) \simeq \pi^{-1}\int_0^{\infty}\Big\{&\left[|W(j\omega)| + \frac{1}{2}\mu\sum_{i=1}^{N}\Delta_i A_i(\omega, t)\right]\\ &\cos(\omega t+\varphi+\alpha) + \frac{1}{2}\mu\sum_{i=1}^{N}\Delta_i B_i(\omega, t)\\ &\sin(\omega t+\varphi+\alpha)\Big\}|\theta_T(j\omega, t)|\,\mathrm{d}\omega\end{aligned} \tag{27}$$

where

$$A_i(\omega, t) = a_i(\omega)\cos\Omega_i t + b_i(\omega)\sin\Omega_i t$$

$$\begin{aligned}a_i(\omega) &= \frac{\partial|W(j\omega)|}{\partial x_i}\left(M_i^+(\omega)\cos\varphi_i^+(\omega) + M_i^-(\omega)\cos\varphi_i^-(\omega)\right)\\ &-|W(j\omega)|\frac{\partial\varphi}{\partial x_i}\left(M_i^+(\omega)\sin\varphi_i^+(\omega) + M_i^-(\omega)\sin\varphi_i^-(\omega)\right)\\ b_i(\omega) &= \frac{\partial|W(j\omega)|}{\partial x_i}\left(M_i^-(\omega)\sin\varphi_i^-(\omega) - M_i^+(\omega)\sin\varphi_i^+(\omega)\right)\\ &-|W(j\omega)|\frac{\partial\varphi}{\partial x_i}\left(M_i^+(\omega)\cos\varphi_i^+(\omega) - M_i^-(\omega)\cos\varphi_i^-(\omega)\right)\end{aligned} \tag{28}$$

$$\begin{aligned}\frac{Q_2(j\omega)}{Q_2\{j(\omega-\Omega_i)\}} &= M_i^-(\omega)\,\mathrm{e}^{j\varphi_i^-(\omega)}\\ \frac{Q_2(j\omega)}{Q_2\{j(\omega+\Omega_i)\}} &= M_i^+(\omega)\,\mathrm{e}^{j\varphi_i^+(\omega)}\end{aligned} \tag{29}$$

[the expressions defining the coefficients $B_i(\omega, t)$ are analogous to formulae (28) and (29)]. In quasi-stationary self-adjustment modes, when Ω_N is small compared with the actual frequencies ω of the test signal $\theta(t)$, $Q_2(j\omega) \simeq Q_2(j(\omega \pm \Omega_i))$, so that

$$M_i^-\,\mathrm{e}^{j\varphi_i^-} \simeq 1 \simeq M_i^+\,\mathrm{e}^{j\varphi_i^+} \tag{30}$$

$$a_i(\omega) \simeq 2\frac{\partial|W(j\omega)|}{\partial x_i}, \qquad b_i(\omega) \simeq 0 \tag{31}$$

§ 7. Most often the detector $\mathscr{D}$ may be considered as either square-law:

$$\varphi(\varepsilon)=\varepsilon^2 \tag{32}$$

or linear:

$$\varphi(\varepsilon)=|\varepsilon|=(\varepsilon^2)^{\frac{1}{2}} \tag{33}$$

In both cases the theoretical analysis requires the square of the error $\varepsilon(t)$ to be calculated. Taking only terms of zero- and first-order smallness with respect to quantity μ, the following expression is readily derived from (27):

$$\varepsilon^2(t)\simeq\frac{1}{2\pi^2}\int_0^\infty\int_0^\infty D(\omega,\nu)\left[\cos((\omega-\nu)t+\vartheta_\omega-\vartheta_\nu)+\cos((\omega+\nu)t+\vartheta_\omega+\vartheta_\nu)\right]\mathrm{d}\omega\,\mathrm{d}\nu \tag{34}$$

where

$$D(\omega,\nu)\geq|\theta_T(j\omega,t)|\,|\theta_T(j\nu,t)|\,|W(j\omega)|\,|W(j\nu)|\times\left[1+\frac{\mu}{2}\sum_{i=1}^{N}\Delta_i\{A_i(\omega,t)\,|W(j\omega)|^{-1}+A_i(\nu,t)\,|W(j\nu)|^{-1}\}\right] \tag{35}$$

while

$$\vartheta_\omega=\varphi(\omega)+\alpha(\omega)+\psi(\omega) \tag{36}$$

$$\tan\psi(\omega)=\frac{\mu}{2}\sum_{i=1}^{N}\Delta_i B_i(\omega,t)\left[|W(j\omega)|+\frac{\mu}{2}\sum_{l=1}^{N}\Delta_l A_l(\omega,t)\right].$$

In all the working below, harmonic search-modulation signals are, in fact, considered and as a mathematical model of System A one takes the system of eqns (19), (32)–(36). These equations continue to hold adequately until the instant when through the operation of the self-adjustment circuits the relation

$$|W(j\omega)|\simeq\mu N\max_{i,\,t}\Delta_i\,|A_i(\omega,t)| \tag{37}$$

becomes true (in that event the approximate expressions in (35) for $\mathscr{D}(\omega,\ \nu)$ are already invalid).

III. Theoretical Analysis of Self-adjustment Modes with an Almost Periodic Test Signal

§ 8. If over the whole time interval that this paper is concerned with the test signal may be represented accurately enough in the form

$$\theta(t)=\theta_0+\sum_{k=1}^{M}\theta_k\cos(\omega_k t+\alpha_k)$$

$$\theta_k,\alpha_k,\omega_k=\text{const},\qquad \omega_k<\omega_{k+1} \tag{38}$$

then it is convenient to consider (38) as being true for $-\infty<t<+\infty$. Then

$$|\theta_T(j\omega,t)|=\pi\sum_{k=0}^{M}\theta_k\delta(\omega-\omega_k),\qquad \omega_0=0 \tag{39}$$

and in accordance with (32)–(35)

$$\varepsilon^2(t)=\frac{1}{2}\sum_{k,l=0}^{M}\theta_k\theta_l\,|W(j\omega_k)|\,|W(j\omega_l)|\,e_{kl} \tag{40}$$

$$e_{kl}=\left[1+\frac{\mu}{2}\sum^{N}\Delta_i A_i(\omega_k,t)\,|W(j\omega_k)|^{-1}+A_i(\omega_l,t)\,|W(j\omega_l)|^{-1}\right]\times\left[\cos((\omega_k-\omega_l)t+\vartheta_k-\vartheta_l)+\cos((\omega_k+\omega_l)t+\vartheta_k+\vartheta_l)\right] \tag{41}$$

where ϑ_k and $A_i(\omega_k,\ t)$ are as defined by (36) and (28).

It can be seen from (41) and (28) that the signal e_{kl} is made up of a sum of harmonics at frequencies $\omega_k\pm\omega_l$, $\omega_k\pm\omega_l\pm\Omega_s$ ($k=1,\ldots,M$, $s=1,\ldots,N$). In the phase discriminator of the ith self-adjustment channel the output quantity $\varphi(\varepsilon)$ of the detector $\mathscr{D}$ is multiplied by the harmonic reference voltage at frequency Ω_i, so that with square-law detection a signal $u_i(t)$ is obtained consisting of harmonics at frequencies $\omega_k\pm\omega_l\pm\Omega_i$, $\omega_k\pm\omega_l\pm\Omega_s\pm\Omega_i$ (with linear detection in general one also gets other harmonic components with amplitudes that are first-order of smallness with respect to quantity μ).

If in (38) θ_0 represents a slowly varying useful signal, while the sum

$$\sum_{k=1}^{M}\theta_k\cos(\omega_k t+\alpha_k) \tag{42}$$

represents intense disturbances at sufficiently high frequencies ($\omega_1>2\,\Omega_N$), then correctly chosen smoothing filters $W_{\phi i}(D)\cdot D^{-1}$ should pass only harmonics of the signal $u_i(t)$ with frequencies

$$\omega_k-\omega_l-\Omega_i,\quad \omega_k-\omega_l-(\Omega_s\pm\Omega_i)\qquad k\geq l \tag{43}$$

It is assumed that the disturbances (42) acting on the system are such that for $k>l$ the conditions

$$\omega_k-\omega_l\neq\Omega_i,\quad \omega_k-\omega_l\neq(\Omega_s\pm\Omega_i) \tag{44}$$

are satisfied with adequate margin. Then it may be considered that the constant component in the signal $u_i(t)$ that is passed by the filter $W_{\phi i}(D)\cdot D^{-1}$ accurately matches that harmonic of the sequence $\omega_k-\omega_l-(\Omega_s-\Omega_i)$ for which $k=l$ and $s=i$. Then according to (41) and the circuit of System A

$$\frac{\mathrm{d}x_i}{\mathrm{d}t}\simeq W_{\varphi i}(D)\cdot E_t(\varepsilon^2(\tau)\,m_{ic}\cos\Omega_i t) \tag{45}$$

where

$$E_t[f(\tau)]=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}f(\tau)\,\mathrm{d}\tau$$

and the values of the parameters X, Y_1 and Y_2 in the expression for $\varepsilon^2(\tau)$ are taken as 'frozen' at the instant t [see (6)], so that

$$E_t[\varepsilon^2(\tau)\,m_{ic}\cos\Omega_i t]=\frac{1}{4}\mu\Delta_i\,m_{ic}\,[E_{ci}^{(1)}+E_{ci}^{(2)}+E_{ci}^{(3)}] \tag{46}$$

$$E_{ci}^{(1)}=\frac{\partial}{\partial x_i}\sum_{k=1}^{M}\theta_k^2\,|W(j\omega_k)|^2\,(M_{ik}^{+}\cos\phi_{ik}^{+}+M_{ik}^{-}\cos\varphi_{ik}^{-}) \tag{47}$$

$$E_{ci}^{(2)}=-\sum_{k=1}^{M}\theta_k^2\,|W(j\omega_k)|^2\frac{\partial}{\partial x_i}(M_{ik}^{+}\cos\varphi_{ik}^{+}+M_{ik}^{-}\cos\varphi_{ik}^{-}) \tag{48}$$

$$E_{ci}^{(3)}=-\sum_{k=1}^{M}\theta_k^2\,|W(j\omega_k)|^2\frac{\partial\varphi(\omega_k)}{\partial x_i}(M_{ik}^{+}\sin\varphi_{ik}^{+}+M_{ik}^{-}\sin\varphi_{ik}^{-}) \tag{49}$$

where the quantities $M_{ik}^{+}\,\mathrm{e}^{j\varphi_{ik}}$ are defined by formulae (29) with $\omega=\omega_k$.

§ 9. The case of the quasi-stationary mode of operation ($\theta_0 = 0$ and condition (30) satisfied) are first considered. Equation (45) for the self-adjustment process becomes

$$\frac{dx_i}{dt} \simeq W_{\varphi i}(D) \cdot \frac{1}{2} \mu \Delta_i m_{ic} \frac{\partial}{\partial x_i} \sum_{k=1}^{N} \theta_k^2 \cdot |W(j\omega_k)|^2 \qquad (50)$$

and thus as a result of the normal operation of the self-adjusting circuit (without loss of stability, without intense distortion caused by disturbances etc.) the quantity

$$\sum_{k=1}^{M} \theta_k^2 |W_2(j\omega_k) - W_1(j\omega_k)|^2 \qquad (51)$$

will be a minimum, i.e. in the complex plane the frequency characteristic of the filter being adjusted will approach that of the calibration filter at the points $\omega = \omega_k$ $(k = 1, \ldots, M)$ in some mean-square sense. If by varying the adjusted parameters X the frequency characteristics $W_1(j\omega)$ and $W_2(j\omega)$ can be made practically identical over some range of frequencies, then this approach will merely signify that over the given frequency range $W_1(j\omega) \simeq W_2(j\omega)$, and the result of the normal operation of the self-adjusting circuit will prove practically independent of the actual spectral composition of the test signal $\sum_{k=1}^{M} \theta_k \cos(\omega_k t + \alpha_k)$, $(\omega_1 \gg \Omega_N)$ (see *Example 1*). The latter statement is not valid (see Taylor[5], and also *Example 2*) if the filters $W_1(j\omega)$ and $W_2(j\omega)$ essentially cannot be made identical. In this event the closest convergence of the frequency characteristics $W_1(j\omega)$ and $W_2(j\omega)$ takes place at those points $\omega = \omega_k$ corresponding to large amplitutes θ_k, and the nature of this convergence will change with variation both of the frequencies ω_k and of the ratios between the amplitudes θ_k.

Example 1. In a System A with square-law detector, let filter W_1 be a controlled plant with transfer function $W_1(p) = (b_2 p + b_1)^{-1}$, and filter W_2 selflearning model[6] with a transfer function of the form $W_2(p) = (X_2 p + X_1)^{-1}$, where X_1 and X_2 are the adjustable parameters, by varying which a complete identity between the dynamic properties of model and plant can, in principle, be achieved. If

$$\theta(t) = \theta \cos(\Omega t + \alpha), \quad \Omega \gg \Omega_2 > \Omega_1 \qquad (52)$$

then eqn (50) for the quasi-stationary self-adjustment mode takes the following form:

$$\dot{X}_1 = k_1 W_{\varphi 1}(D) \cdot [(X_1 - b_1)(X_2^2 \Omega^2 + X_1 b_1) - \Omega^2 (X_2 - b_2)^2 X_1] \qquad (53)$$

$$\dot{X}_2 = k_2 W_{\varphi 2}(D) \cdot [(X_2 - b_2)(X_1^2 + X_2 b_2 \Omega^2) - X_2 (b_1 - x_1)^2] \qquad (54)$$

$$k_i = \mu \Delta_i m_{ic} \theta^2 (b_1^2 + b_2^2 \Omega^2)^{-1} \cdot (X_1^2 + X_2^2 \Omega^2)^{-2}, \quad i = 1, 2 \qquad (55)$$

It can be seen from eqns (53)–(55) that as a result of the normal operation of the self-adjusting circuit $X_1 \to b_1$ and $X_2 \to b_2$, i.e. in fact $W_2(j\omega) \to W_1(j\omega)$, at whatever frequency Ω the test signal (52) is applied. The self-adjustment process forms a coupled control of the parameters X_1 and X_2. The higher the frequency Ω of the test signal, the more intensively the adjustment of X_2 takes place (*cf.* Margolis and Leondes[6]). The stability for small variations of the equilibrium $X_1 = b_1$, $X_2 = b_2$ in the non-linear system (53)–(55) can readily be examined from the first-order approximation equations.

Example 2. If in the system just considered the controlled plant is close in its dynamic properties to the link $W_1(p) = e^{-p\tau}(b_2 p + b_1)^{-1}$, while $W_2(p) = (X_2 p + X_1)^{-1}$, then for $\tau \neq 0$ complete identity of filters W_2 and W_1 cannot be achieved by self-adjustment. Transcribing eqn (50) for this process, it can easily be seen that the result

$$X_1 \to b_1 \cos \Omega\tau + b_2 \Omega \sin \Omega\tau,$$

$$X_2 \to b_2 \cos \Omega\tau - b_1 \Omega^{-1} \sin \Omega\tau$$

of normal operation of the self-adjustment network may already depend substantially on the frequency of the test signal (52).

§ 10. Eqns (45)–(49) also permit a number of conclusions of a qualitative nature about non-quasi-stationary modes of self-adjustment in system A ($\theta^0 \neq 0$, conditions (30) not satisfied) to be immediately drawn.

It is first observed that the equation

$$Q_2 = Q_2(j\omega) \qquad (56)$$

defines a Mikhaylov hodograph[10] for a stable [assumption (*b*)] linear system, and consequently the curve (56) has a form similar to that in *Figure 2*.

It then becomes clear from (45)–(49) that within the limits of the errors introduced by the terms $E_{ci}^{(2)}$ and $E_{ci}^{(3)}$ the normal operation of the ith self-adjustment channel reduces to the minimization of the quantity

$$\sum_{k=0}^{M} \theta_k^2 |W(j\omega_k)|^2 (M_{ik}^{+} \cos \varphi_{ik}^{+} + M_{ik}^{-} \cos \varphi_{ik}^{-}) \qquad (57)$$

The self-adjustment error associated with $E_{ci}^{(2)}$ will be small in most cases, since by the very sense of the quantities $M_{ik}^{\pm} e^{j\varphi_{ik}^{\pm}}$ [see (29) and also *Figure 2*] the partial derivatives $\partial/\partial x_i$ $(M_{ik}^{+} \cos \varphi_{ik}^{+} + M_{ik}^{-} \cos \varphi_{ik}^{-})$ will hardly be significantly different from zero. The error associated with $E_{ci}^{(3)}$ will also be insignificant, since with $|\varphi_{ik}^{+}|$, $|\varphi_{ik}^{-}| < \pi$ the terms $M_{ik}^{+} \sin \varphi_{ik}^{+}$ and $M_{ik}^{-} \sin \varphi_{ik}^{-}$ are opposite in sign.

If

$$M_{ik}^{+} \cos \varphi_{ik}^{+} + M_{ik}^{-} \cos \varphi_{ik}^{-} > 0, \quad k = 1, \ldots, M \qquad (58)$$

then the minimization of the quantity (57) has roughly the same physical significance (see § 9) as the minimization of quantity (51) in quasi-stationary modes, and thus the result of normal operation of the self-adjustment network should be taken as acceptable. But the more strongly the self-adjustment mode differs from quasi-stationary, the larger are the angles $|\varphi_{ik}^{\pm}|$ and $|\varphi_{ik}^{\pm}| < \pi$ the smaller are for the coefficients in (58) (in particular, the quantity $M_{i1}^{+} \cos \varphi_{i1}^{+} + M_{i1}^{-} \cos \varphi_{i1}^{-}$), since over a substantial range of frequencies the quantities $M_{ik}^{\pm}$ are hardly much different from unity. As a result the quality factor for the X_i tracking system falls, while for $|\varphi_{ik}^{\pm}| > \pi/2$ the coefficient (58) becomes negative and minimization of the weighted sum (57) of squares $|W_2(j\omega_k) - W_1(j\omega_k)|^2$ loses its evident sense, or even on inversion of the self-adjusting servo-system occurs (particularly if all the coefficients (58) become negative, which

may happen if a strong signal $\theta(t)$ is applied at a frequency close to a search frequency Ω_i—see *Example 3*).

Example 3. Let the adjustable filter in System A be a link with transfer function $W_2(p) = kQ_2(p)^{-1} = k(p^2 + 2\alpha p + \omega_0^2)^{-1}$, the adjustment parameter being the gain k, modulated by a signal $\Delta k \cdot \cos \Omega t$, while to the input of the system is applied the test action $\theta(t) = \theta \cos(\omega t + \alpha)$.

Writing out the general expressions for the quantities

$$M^{\pm} e^{j\varphi^+} = Q_2(j\omega) \cdot Q_2[j(\omega \pm \Omega)]^{-1} \tag{59}$$

it can readily be established that condition (58) for the system considered is explicitly inobserved if ω_0 is small ($\omega_0 \to 0$), while the frequencies Ω and ω of the search and test signals coincide and exceed ω_0, since then

$$M^+ \cdot \cos\varphi^+ \to \frac{1}{4}(\Omega^2 + 2\alpha^2)(\Omega^2 + \alpha^2)^{-1}$$

$$M^- \cos\varphi^- \to -\Omega^2\omega_0^2$$

$$M^+ \sin\varphi^+ \to -\frac{1}{4}\alpha(\Omega^2 + \alpha^2)^{-1}$$

$$M^- \sin\varphi^- \to 2\alpha\omega_0^{-2}$$

and the coefficient $M^- \cos\varphi^-$ becomes in (58) greater in modulus than the coefficient $M^+ \cos\varphi^+$.

It is observed that since in this case the quantities (59) are independent of k, the error associated with the term $E_{ci}^{(2)}$ in eqns (45)–(49) proves equal to zero [this situation will occur evry time that the adjustable parameters of filter ϕ_2 appear only in the numerator of the transfer function $W_2(p)$].

Considering eqns (40), (41) and (28), it is noted that to increase the capability of the self-adjusting circuit for operating in non-quasi-stationary conditions one may use in the phase discriminators ϕD_i the reference voltages

$$m_{ic} \cos \Omega_i t + m_{is} \sin \Omega_i t \tag{60}$$

which are phase shifted with respect the search modulation signal

$$\mu\Delta_i \cos \Omega_i t \tag{61}$$

In this case the processes of self-adjustment will proceed in accordance with the equations

$$\dot{X} \simeq W_{\varphi i}(D)[E_t(\varepsilon^2(t)\, m_{ic} \cos \Omega_i t) + E_t(\varepsilon^2(t)\, m_{is} \sin \Omega_i(t)] \tag{62}$$

where $E_t(\varepsilon^2(t)\, m_{ie} \cos \Omega_i t)$ is determined by formulae (46)–(49), while

$$E_t(\varepsilon^2(t)\, m_{is} \sin \Omega_i t) = \frac{1}{4}\mu\Delta_i\, m_{is}(E_{si}^{(1)} + E_{si}^{(2)} + E_{si}^{(3)}) \tag{63}$$

$$E_{si}^{(1)} = \frac{\partial}{\partial x_i} \sum_{k=0}^{M} \theta_k^2 |W(j\omega_k)|^2 (M_{ik}^- \sin\varphi_{ik}^- - M_{ik}^+ \sin\varphi_{ik}^+) \tag{64}$$

$$E_{si}^{(2)} = -\sum_{k=0}^{M} \theta_k^2 |W(j\omega_k)|^2 \frac{\partial}{\partial x_i}(M_{ik}^- \sin\varphi_{ik}^- - M_{ik}^+ \sin\varphi_{ik}^+) \tag{65}$$

$$E_{si}^{(3)} = -\sum_{k=0}^{M} \theta_k^2 |W(j\omega_k)|^2 \frac{\partial\varphi(\omega_k)}{\partial x_i}(M_{ik}^+ \cos\varphi_{ik}^+ - M_{ik}^- \cos\varphi_{ik}^-) \tag{66}$$

Here the necessary condition (58) for normal operation of the self-adjusting circuits is replaced by the condition

$$M_{ik}^+(m_{ic}\cos\varphi_{ik}^+ - m_{is}\sin\varphi_{ik}^+) + M_{ik}^-(m_{ic}\cos\varphi_{ik}^- + m_{is}\sin\varphi_{ik}^-) > 0 \tag{67}$$

which may prove much more favourable given a suitable choice of the phase of the voltage (60); (i.e. of the quantities m_{ic} and m_{is}) the actual result of the undistorted forced process of self-adjustment comes out in this case to be the minimization of the quantity

$$\sum_{k=0}^{M} \theta_k^2 |W(j\omega_k)|^2 [M_{ik}^+(m_{ic}\cos\varphi_{ik}^+ - m_{is}\sin\varphi_{ik}^+) + M_{ik}^-(m_{ic}\cos\varphi_{ik}^- + m_{is}\sin\varphi_{ik}^-) \tag{68}$$

In choosing the phase of the reference voltage (60) one can aim not only at increasing the coefficient (67) but also at the same time decreasing the quantity

$$|M_{ik}^+(m_{ic}\sin\varphi_{ik}^+ + m_{is}\cos\varphi_{ik}^+) + M_{ik}^-(m_{ic}\sin\varphi_{ik}^- + m_{is}\cos\varphi_{ik}^-)| \tag{69}$$

i.e. (see (62), (66) and (49)) the error associated with the term $E_{ci}^{(3)}$. In practice, as a rule, it proves tedious to achieve an accurately optimum phase-shift (e.g. in the sense of a minimum ratio between the quantities (69) and (67) between the signals (60) and (61), since by virtue of (29) this shift depends not only on the drifting parameters of filter ϕ_2 (a similar situation arises[11] also in extremal control systems), but also on the form of the test signal $\theta(t)$. Nevertheless by using *a priori* information on the operating conditions of the system, or by carrying out a running analysis of the signal $\theta(t)$ and the results of system operation, in a number of cases one can evidently achieve an improvement in the dynamic properties of the given self-adjusting system relatively simply by using reference voltages of the form in (60) that only approximate to the optimum. In order to increase the stability of automatic phase-shift optimization between voltages (60) and (61) one can correlate the search and test signals in frequency [phase relations between the signals $\theta(t)$ and $\mu\Delta_i \cos \Omega_i t$ have no effect on the quantities (47)–(49), (64)–(66)].

The self-adjusting system, the phase of which use discriminators reference voltages of the general type given in (60) will be denoted by System B.

§ 11. The equations of motion (45)–(49) and (62)–(66) were derived under the assumption that the frequencies of the search modulation and the harmonic components of the test signal all satisfy the conditions (44). If these conditions do not hold, then the voltages $Et(\varepsilon^2(t)\, m_{ic} \cos \Omega_i t)$ and $E_t(\varepsilon^2(t)\, m_{is} \sin\Omega_i t)$, together with the signals (47)–(49) and (64)–(66), will also contain other components, which generally speaking will introduce certain additional distortions into the self-adjustment process. Equations (40) and (41) enable one effectively to calculate all these parasitic components of the control signal in the self-adjusting network.

For example, let only one of the conditions (44) be disturbed: let the frequency of the pth harmonic of the test signal coincide with the search frequency in the ith self-adjusting channel, i.e. $\omega p = \Omega_i$. According to (41), in this case the signal $E_t(\varepsilon^2(t)\, m_{ic} \cos \Omega_i t)$ will contain an additional term $E_{ci}^{(4)}$,

generated by the presence in e_{kl} of harmonics with frequencies $\omega_k + \omega_l - \Omega_q$ (for $k = 0, l = p, q = i$ and $k = p, l = 0, q = i$) and $\omega_k + \omega_l - (\Omega_s + \Omega_q)$ (for $k = l = p, s = q = i$):

$$E_{ci}^{(4)} = \frac{1}{2} m_{ic} E_t [\theta_0 \theta_p |W(0) W(j\omega_p)| (e_{po} + e_{op}) + \frac{1}{2} \theta_p^2 |W(j\omega_p|^2 e_{pp}] \cos \Omega_i t = m_{ic} \theta_0 \theta_p |W(0) W(j\omega_p)| \cos \vartheta_p + \frac{1}{4} \mu \Delta_i m_{ic} \theta_p^2 |W(j\omega_p)| x [a_i(\omega_p) \cos 2\vartheta_p - b_i(\omega_p) \sin 2\vartheta_p] \quad (70)$$

where $\vartheta_q = \vartheta(\omega_q)$, $a_i(\omega_q)$ and $b_i(\omega_p)$ are defined respectively by eqns (36), (28) and (29) with $\omega = \omega_p$.

For the system considered in *Example 3*, the first term in expression (10) is zero (since $\theta_0 = 0$), while the second may be calculated given the frequency characteristic of $W_1(j\omega)$. Even in this actual example it is, on the whole, difficult to judge what effect the use of a reference voltage of (60) type will have on the additional error in question. One can evidently achieve a stable reduction in this error or even its conversion into a useful signal, provided one correlates the search and test signals not only in frequency but also in phase, so as to limit unforeseen variations in the angle ϑ_p.

IV. Calculation of Self-adjustment Operating Modes where the Test Action is a Stationary Random Process

§ 12. It is assumed for simplicity that the filters $W_{\phi i}(D)$ in System A consist of elements which carry out the ideal averaging of the quantity $m_{ic}\, \varepsilon^3(t) \cos \Omega_i t$ in time over the interval $(t - T_0, t)$:

$$W_{\varphi i}(D)[\varepsilon^2(t)\, m_{ic} \cos \Omega_i t] = \frac{k_i}{T_0} \int_{t-T_0}^{t} \varepsilon^2(\tau) \cos \Omega_i \tau \, d\tau \quad (71)$$

and that the test signal $\theta(t)$ is a time-function whose 'shortened' spectrum (10) actual only slightly depends on the instant of observation t and is located in the region of quite high frequencies:

$$\theta_T(j\omega, t) \simeq \theta_T(j\omega), \quad \theta_T(j\omega) = 0 \text{ for } \omega < \omega^* \quad (72)$$

$$\omega^* - 2\Omega_N \geq T_0^{-1}, \quad \Omega_i - \Omega_{i-1} \geqslant T_0^{-1} \quad (73)$$
$$(\Omega_i > \Omega_{i-1}, \quad i = 1, \ldots, N)$$

Every actual filter $W(j\omega) = W_1(j\omega) - W_2(j\omega)$ has a finite cut-off frequency ω_φ (it is further considered that $\omega^* < \omega_\varphi$), so that in accordance with (19), (32), (34) and (71)–(73) the equations for the process of self-adjustment of the qth parameter may be put into the form

$$\dot{X}_q \simeq \pi^{-2} \int_{\omega^*}^{\omega_\varphi} d\omega \int_{\omega}^{\omega_\varphi} d\nu T_0^{-1} \int_{t-T}^{t} G_q(\omega, \nu, \tau) \, d\tau \quad (74)$$

$$G_q(\omega, \nu, \tau) = D(\omega, \nu) [\cos \{(\omega - \nu)\tau + \vartheta_\omega - \vartheta_\nu\} + \cos \{(\omega + \nu)\tau + \vartheta_\omega + \vartheta_\nu\}] \cos \Omega_q \tau$$

where $D(\omega, \nu)$, ϑ_ω and $\vartheta\nu$ are defined by eqns (35), (36), (28) and (29). The quantity $G_q(\omega, \nu, t)$ is a sum of harmonic components with frequencies Ω equal to

$$\omega \pm \nu \pm \Omega_q, \quad \omega \pm \nu \pm (\Omega_s \pm \Omega_q), \quad (s = 1, \ldots, N) \quad (75)$$

while the integral

$$\int_{t-T_0}^{t} G_q(\omega, \nu, \tau) \, d\tau$$

is a weighted sum of integrals of the type

$$\int_{t-T_0}^{t} \cos(\Omega\tau + \vartheta) \, d\tau \quad (76)$$

where Ω are the frequencies in (75) and ϑ are angles of the form $\vartheta_\omega \pm \vartheta\nu$ and $\vartheta_\omega \pm \vartheta_\nu + \pi/2$. On rewriting the integral (76) in the form

$$\int_{-\frac{1}{2}T_0}^{\frac{1}{2}T_0} \cos\left[\Omega\left(\xi + t - \frac{1}{2}T_0\right) + \vartheta\right] d\xi$$

it is observed that in accordance with a known[8] integral representation

$$\frac{1}{2}\pi^{-1} \int_{-\infty}^{\infty} \cos(\Omega\tau + \vartheta) \, d\tau = \cos\vartheta \cdot \delta(\Omega) \quad (77)$$

of the δ function and for large enough averaging time intervals T_0 of the filter $W_{\phi q}(D)$, the approximate equation

$$\int_{t-T_0}^{t} \cos(\Omega\tau + \vartheta) \, d\tau \simeq \cos\left[\Omega\left(t - \frac{1}{2}T_0\right) + \vartheta\right] \delta(\Omega) = \cos\vartheta \cdot \delta(\Omega) \quad (78)$$

is true; using this, eqn (74) can be readily got into the form

$$\dot{X}_q \simeq T \cdot T_0^{-1} \cdot k_q \cdot \pi^{-1} \int_{\omega^*}^{\omega_\varphi} G_q(\omega) + \frac{\mu}{2} \sum_{i=1}^{N} \Delta_i [g_i(\omega, \nu_{iq}^+) + g_i(\omega, \nu_{iq}^-) + g_q(\omega, \nu_{qq}^-)] \, d\omega \quad (79)$$

where

$$G_q(\omega) = |\theta_T(j\omega)\, \theta_T\{j(\omega + \Omega_q)\}\, W(j\omega)\, W\{j(\omega + \Omega_q)\}| \cdot T^{-1} \cos(\vartheta_\omega - \vartheta_{\omega+\Omega_q}) \quad (80)$$

$$g_i(\omega, \nu) = \frac{1}{2} T^{-1} |\theta_T(j\omega)\, \theta_T(j\nu)\, W(j\omega)\, W(j\nu)| \cdot [V_{ic}(\omega, \nu) \cos(\vartheta_\omega - \vartheta_\nu) - V_{is}(\omega, \nu) \sin(\vartheta_\omega - \vartheta_\nu)] \quad (81)$$

$$V_{ic}(\omega, \nu) = a_i(\omega) |W(j\omega)|^{-1} + a_i(\nu) |W(j\nu)|^{-1}$$
$$V_{is}(\omega, \nu) = b_i(\omega) |W(j\omega)|^{-1} + b_i(\nu) |W(j\nu)|^{-1} \quad (82)$$

$$\nu_{iq}^+ = \omega + \Omega_i + \Omega_q, \qquad \nu_{iq}^- = \omega + |\Omega_i - \Omega_q| \quad (83)$$

[the quantities $a_i(\omega)$, $b_i(\omega)$ and ϑ_ω being defined by eqns (28), (29) and (36) and l the memory of filter $W(j\omega)$—see § 3].

Considering the function (5) as a typical realization of a stationary random process $\{\theta(t)\}$ and performing averaging according to achievements, one can go from eqns (79)–(83) to equations in the mean (as taken together) values $\bar{X}_q$ of the adjustable parameters. If here the interval T is taken large enough, then in the right-hand sides of these equations one may replace the quantities $T^{-1}\overline{\theta_T(j\omega)\, \theta_T(j\nu)}$ by characteristics like the mutual spectral power densities[12] of the process $\{\theta(t)\}$ and certain random processes obtained from $\{\theta(t)\}$ by simple transformations that do not infringe the stationary condition.

This paper does not deal with the more detailed analysis of the general case, but gives the results of the calculation for the quasi-stationary mode of self-adjustment, i.e. the mode in which

$$\omega^* > 2\,\Omega_N \quad (\Omega_i > \Omega_{i-1},\ i = 1, \ldots, N) \tag{84}$$

with a test signal of white-noise type:

$$T^{-1}\,\overline{|\theta_T(j\omega)|^2} \simeq \lim_{T\to\infty} T^{-1}\,\overline{|\theta_T(j\omega)|^2} = \begin{cases} 0 & \text{for } \omega < \omega^* \\ G_\theta & \text{for } \omega > \omega^* \end{cases} \tag{85}$$

Since eqns (30) and (31) are satisfied in quasi-stationary modes, and furthermore $\vartheta_\omega \simeq \vartheta_{\omega+2\Omega M}$ $(\omega > \omega^*)$, one may neglect the terms $V_{is}(\omega, \nu) \sin(\vartheta_\omega - \vartheta_\nu)$ in (81), and so putting $\theta_T(j\omega) \simeq \theta_T\{j(\omega + 2\Omega_M)\}$ and $W(j\omega) \simeq W\{j(\omega + 2\Omega_M)$, the following equations for the self-adjustment process are arrived at:

$$\dot{\bar{X}}_q \simeq k_q^0 \left[\int_{\omega^*}^{\omega_\varphi} |W(j\omega)|^2\,\mathrm{d}\omega + \mu\Delta_q \frac{\partial}{\partial x_q} \int_{\omega^*}^{\omega_\varphi} |W(j\omega)|^2\,\mathrm{d}\omega + \frac{1}{2}\mu \sum_{\substack{i=1 \\ i\neq q}}^{N} \Delta_i \frac{\partial}{\partial x_i} \int_{\omega^*}^{\omega_\varphi} |W(j\omega)|^2\,\mathrm{d}\omega \right] \tag{86}$$

$$k_q^0 = T \cdot T_0^{-1} \cdot k_q^{\omega^*} \cdot \frac{1}{\pi} \cdot m_{qc} \cdot G_\theta$$

The following conclusions are evident from (86):

(1) In the mode of operation (84), (85) studied, minimization of the quantity

$$\int_{\omega^*}^{\omega_\varphi} |W(j\omega)|^2\,\mathrm{d}\omega \tag{87}$$

may be naturally considered the ideal result of the self-adjustment process.

(2) The control signal for the qth self-adjusting network contains derivatives of the quantity (87) being minimized, not only w.r.t. X_q but also w.r.t. all the other adjustable parameters X_i, so that one has not got a pure gradient system of extremal control.

(3) The equilibrium condition $\dot{\bar{X}}_q = 0$ $(q = 1, \ldots, N)$ for the system (86) is characterized for $\Delta_i = \Delta$ $(i = 1, \ldots, N)$ by the relations

$$\mu\Delta(N+1)\frac{\partial}{\partial x_i}\int_{\omega^*}^{\omega_\varphi} |W(j\omega)|^2\,\mathrm{d}\omega = -\int_{\omega^*}^{\omega_\varphi} |W(j\omega)|^2\,\mathrm{d}\omega \quad (i=1, \ldots, N) \tag{88}$$

from which it can be seen that the more pronounced the extremal nature of the dependence of quantity (87) on the parameters X_i, and the less essentially attainable the minimum of this quantity, the closer will this condition be to the ideal result of self-adjustment.

(4) If quite large differences arise rapidly between the frequency characteristics $W_1(j\omega)$ and $W_2(j\omega)$, the non-negative term (87) on the right-hand side of eqn (86) will increase so much that the operation of the self-adjusting network will be reduced merely to increasing the parameter $\dot{\bar{X}}_q$ $(\dot{\bar{x}}_q > 0)$, and this may lead to the system's losing its required extremal condition.

Finally it is noted that the equations given by Krasovskiy[2] for quasi-stationary self-adjustment with a white-noise test signal contain only terms analogous to the second term in the right-hand side of equation (86).

The author expresses his gratitude to Ye. A. Barbashin and I. N. Pechorina for their discussion of this paper.

References

[1] KRASOVSKIY, A. A. Self-adjusting automatic control systems. *Automatic Control and Computer Engineering*. 1961. No. 4. Mashgiz

[2] KRASOVSKIY, A. A. The dynamics of continuous automatic control systems with extremal self-adjustment of the correcting devices. *Automatic and Remote Control*. 1960. London; Butterworths

[3] KAZAKOV, I. YE. The dynamics of self-adjusting systems with extremal continuous adjustment of the correcting networks in the presence of random perturbations. *Automat. Telemech.* 21, No. 11 (1960)

[4] VARYGIN, V. N. Some problems in the design of systems with extremally self-adjusting correcting devices. *Automat. Telemech.* 22, No. 1 (1961)

[5] TAYLOR, W. K. An experimental control system with continuous automatic optimization. *Automatic and Remote Control*. 1960. London; Butterworths

[6] MARGOLIS, M., and LEONDES, K. T. On the theory of self-adjusting control systems, the learning model method. *Automatic and Remote Control*. 1960. London; Butterworths

[7] ITSKHOKI, YA. S. *Non-Linear Radio Engineering*. 1955. Sovetskoye Radio

[8] KHÄRKEVICH, A. A. *Spectra and Analysis*. 1953. Gostekhizdat

[9] MALKIN, I. G. *Some Problems in the Theory of Non-Linear Oscillation*. 1956. Gostekhizdat

[10] POPOV, YE. P. *The Dynamics of Automatic Control Systems*. 1954. Gostekhizdat

[11] CH'IEN HSÜEH-SEN. *Technical Cybernetics*. 1956. Izd. Inostr. Lit.

[12] LANING, G. H., and BETTIN, R. G. *Random Processes in Automatic Control Problems* (Russian transl.). 1958. Izd. Inostr. Lit.

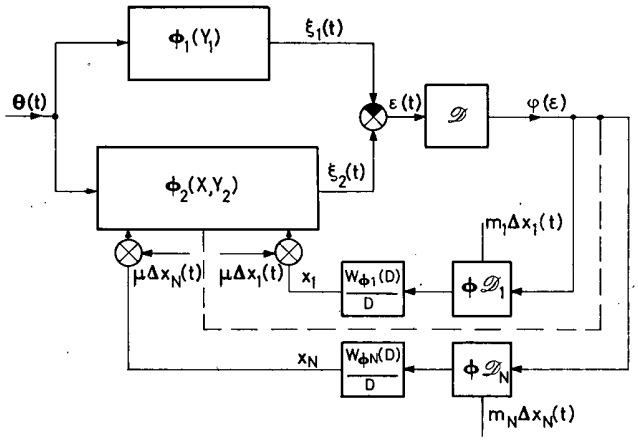


Figure 1

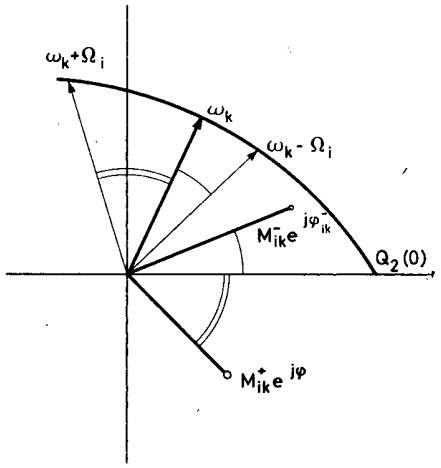


Figure 2 →

Optimal Control of Systems with Distributed Parameters

A. G. BUTKOVSKIY

In many engineering applications the need arises for control of systems with parameters that are distributed in space. A wide class of industrial and non-industrial processes falls within this category: production flow processes, heating of metal in methodical or straight-through furnaces before rolling or during heat-treatment, establishment of given temperature distributions in 'thick' ingots, growing of monocrystals, drying and calcining of powdered materials, sintering, distillation, etc., right through to the control of the weather.

The processes in such systems are normally described by partial differential equations, integral equations, integro-differential equations, etc.

The problem of obtaining the best operating conditions for the installation (the highest productivity, minimum expenditure of raw material and energy, etc.) under given additional constraints has required the development of an appropriate mathematical apparatus capable of determining the optimal control actions for the plant.

Pontryagin's maximum principle and Bellman's dynamic programming method have been the most interesting results in this direction for systems with lumped parameters.

A wide class of systems with distributed parameters is described by a non-linear integral equation of the following form:

$$Q(P)=\int_D K[P,S,Q(S),U(S)]\,dS \tag{1}$$

Here the matrix

$$Q(P)=\left\|\begin{matrix}Q^1(P)\\ \vdots \\ Q^n(P)\end{matrix}\right\|=\|Q^i(P)\| \tag{2}$$

describes the condition of the controlled system with distributed parameters, while the matrix

$$U(P)=\left\|\begin{matrix}U^1(P)\\ \vdots \\ U^r(P)\end{matrix}\right\|=\|U^i(P)\| \tag{3}$$

describes the control actions on the system. Here and in the following, the index i will refer to a row number and j to a column number in a matrix. The point P belongs to a certain fixed m dimensional region D in Euclidean space.

The components of the single-column matrix

$$K(P,S,Q,U)=\left\|\begin{matrix}K^1(P,S,Q,U)\\ \vdots \\ K^n(P,S,Q,U)\end{matrix}\right\|=\|K^i(P,S,Q,U)\| \tag{4}$$

belong to class L_2 and have continuous partial derivatives w.r.t. the components of the matrix Q.

It will be assumed that the function $U(P)$ is piecewise discontinuous, its values being chosen from a certain fixed permissible set Ω. Controls $U(P)$ having this property will be called permissible.

Further, from the set of conditions $Q(P)$ and controls $U(P)$, related by integral eqn (1), let q functionals be determined, having a continuous gradient (weak Gato differential).

$$I^i=I^i[Q(P)],\quad i=0,1,\ldots,l \tag{5}$$

$$I^i=I^i[Q(P),U(P)]=\Phi^i(z),\quad i=l+1,\ldots,q \tag{6}$$

where

$$z=\left\|\begin{matrix}z^0\\ \vdots \\ z^k\end{matrix}\right\|=\left\|\begin{matrix}\int_D F^0[S,Q(S),U(S)]\,dS\\ \vdots \\ \int_D F^k[S,Q(S),U(S)]\,dS\end{matrix}\right\|$$

$$=\left\|\int_D F[S,Q(S),U(S)]\,dS\right\| \tag{7}$$

The function $\Phi^i(z), i=l+1,\ldots,q$ and $F^i(S,Q,U), i=0,1\ldots,k$, are continuous and have continuous partial derivatives w.r.t. the components of the matrices z and Q respectively.

The optimal control problem is formulated in the following manner.

It is required to find a permissible control $U(P)$ such that by virtue of equation (1)

$$I^i=0,\quad i=0,1,\ldots,p-1,p+1,\ldots,q \tag{8}$$

while the functional I^p assumes its smallest value. Here p is a fixed index, $0\le p\le q$.

The following rectangular matrices are introduced

$$\frac{\partial\Phi}{\partial z}=\left\|\frac{\partial\Phi^i}{\partial z^j}\right\|;\quad i=0,1,\ldots,l;\quad j=0,1,\ldots,k \tag{9}$$

$$\frac{\partial F}{\partial Q}=\left\|\frac{\partial F^i}{\partial Q^j}\right\|;\quad i=0,1,\ldots,k;\quad j=1,2,\ldots,n \tag{10}$$

$$\operatorname{grad} I=\|\operatorname{grad}_j I^i\|;\quad i=l+1,\ldots,q;\quad j=1,2,\ldots,n \tag{11}$$

where $\operatorname{grad}_j I^i$ denotes the jth component of the vector grad I^i w.r.t. the coordinate Q^j.

The following theorem[5] can be used as the basis of a solution of the problem formulated above on the optimum control of a plant with distributed parameters.

Theorem. Let $U=U(S)$ be a permissible control such that by virtue of eqn (1) the conditions (8) are satisfied and the

matrix function $M(P, R) = \|M_{ij}(P, R)\|$, $i, j = 1, 2, \ldots, n$, satisfies the integral equation [linear in $M(P, R)$]

$$M(P,R)+\frac{\partial}{\partial Q}K[P,R,Q(R),U(R)]$$
$$=\int_D M(P,S)\frac{\partial}{\partial Q}K[S,R,Q(R),U(R)]\,dS$$
$$=\int_D \frac{\partial}{\partial Q}K[P,S,Q(S),U(S)]\,M(S,R)\,dS \qquad (12)$$

Then for this control, $U(S)$, to be optimal there must exist one-row numerical matrices

$$a=\|c_0, c_1, \ldots, c_l\| \quad \text{and} \quad b=\|c_{l+1}, \ldots, c_q\| \qquad (13)$$

of which at least one is not null, and also $c_p \leq 0$, such that for almost all fixed values of the argument $S \in D$ the function

$$\pi(S,U)=a\,[\operatorname{grad} I\{Q(P)\}, K\{P,S,Q(S),U\}$$
$$-\int_D M(P,R)\,K\{R,S,Q(S),U\}\,dR]$$
$$+b\frac{\partial}{\partial z}\Phi\left[\int_D F\{P,Q(P),U(P)\}\,dP\right]$$
$$\cdot\left[\frac{\partial}{\partial Q}F\{P,Q(P),U(P)\}, K\{P,S,Q(S),U\}\right.$$
$$\left.-\int_D M(P,R)\,K\{R,S,Q(S),U\}\,dR\right]$$
$$+b\frac{\partial}{\partial z}\Phi\left[\int_D F\{P,Q(P),U(P)\}\,dP\right]\cdot F\{S,Q(S),U\} \qquad (14)$$

of the variable $U \in \Omega$ attains a maximum, i.e. for almost all $S \in D$ the following relation holds:

$$\pi(S,U)=H(S) \qquad (15)$$

where

$$H(S)=\sup_{u\in\Omega}\pi(S,U) \qquad (16)$$

As an example of the application of this theorem, consider the important practical problem of the heating of a massive body in a furnace. Let the temperature distribution along the x axis, $0 \leq x \leq L$, at any instant t, $0 \leq t \leq T$, be described by the function $Q = Q(x, t)$. Here the temperature $U(t)$ of the heating medium, which in this case is the controlling agent, is a function constrained by the conditions

$$A_1 \leq U(t) \leq A_2, \quad 0 \leq t \leq T \qquad (17)$$

i.e. in this case the set Ω is the interval $[A_1, A_2]$.

It is known that the distribution function $Q(x, t)$, if initially zero, is related to the control $U(t)$ by the following integral equation

$$Q(x,t)=\int_0^t K(x,t,\tau)\,U(\tau)\,d\tau \qquad (18)$$

where $K(x, t, \tau)$ is a known weighting function.

In the heating of a body there is usually given a temperature distribution $Q^* = Q^*(x)$ which is required to be attained in the minimum time. However, if the equation

$$Q(x,t)=Q^*(x) \qquad (19)$$

for any permissible control is not satisfied for any fixed t, $0 \leq t \leq T$, then the problem becomes that of determining a permissible control $u(t)$, $0 \leq t \leq T$, such that the functional

$$I^0=\int_0^L [Q^*(x)-Q(x,T)]^\gamma\,dx \qquad (20)$$

attains its minimum. Here γ is a positive even integer.

Since the integrand in eqn (18) is independent of the controlled function $Q(x, t)$, then according to eqn (12) the function $M(t, c) \equiv 0$ for all t and τ in the interval $[0, T]$.

It follows that the function $\pi(\tau, U)$ takes the form

$$\pi(\tau,U)=c_0\int_0^L \frac{\partial}{\partial Q}[Q^*(x)-Q]^\gamma\cdot K(x,T,\tau)\,U\,dx$$
$$=-\gamma c_0 U\int_0^L [Q^*(x)-Q(x,T)]^{\gamma-1}K(x,T,\tau)\,dx \qquad (21)$$

Since in this case by the conditions of the theorem $c^0 < 0$, so $-\gamma C_0 > 0$, and hence the maximum of $\pi(\tau, U)$ w.r.t. U, with $A_1 \leq U \leq A_2$, is reached when

$$U(\tau)=\frac{A_1+A_2}{2}$$
$$+\frac{A_2-A_1}{2}\operatorname{sgn}\int_0^L [Q^*(x)-Q(x,T)]^{\gamma-1}K(x,T,\tau)\,dx \qquad (22)$$

If we substitute expression (18) for $Q(x, t)$ in eqn (22), then we obtain an integral equation for determining the optimum control action $U(\tau)$.

For example, if $\gamma = 2$, $A_1 = -1$, $A_2 = 1$, then the optimum control action satisfies the following integral equation:

$$U(\tau)=\operatorname{sgn}\int_0^L\left[Q^*(x)-\int_0^T K(x,T,\tau)\,U(\tau)\,d\tau\right]K(x,T,\tau)\,dx \qquad (23)$$

Opening the brackets and altering the order of integration, one finally gets

$$U(\tau)=\operatorname{sgn}\left[B(\tau)-\int_0^T N(\tau,\theta)\,U(\theta)\,d\theta\right] \qquad (24)$$

where $N(\tau, \theta)$ is the symmetrical nucleus

$$N(\tau,\theta)=\int_0^L K(x,T,\tau)\,K(x,T,\theta)\,dx \qquad (25)$$

$$B(\tau)=\int_0^L Q^*(x)\,K(x,T,\tau)\,dx \qquad (26)$$

Methods of approximating partial differential equations by finite difference equations can be applied successfully to the approximate solution of problems of the optimal control of systems with distributed parameters. This has the advantage

that results obtained for lumped-parameter optimal systems can be used.

As an example, consider the optimal control of a system described by the following equation

$$\frac{\partial Q}{\partial t}=a\frac{\partial^2 Q}{\partial x^2},\quad Q=Q(x,t),0\le x\le S,0\le t\le T \tag{27}$$

with these initial and boundary conditions

$$Q(x,0)=Q_0(x) \tag{28}$$

$$\left.\frac{\partial Q}{\partial x}\right|_{x=0}=\alpha[U(t)-Q(0,t)],\qquad \left.\frac{\partial Q}{\partial x}\right|=0 \tag{29}$$

Also let the function $Q^*=Q^*(x)$ be given. The problem may be formulated in double form:

(*a*) To find a permissible control $U(t)$, $0\le t\le T$, $U\in\Omega$ (Ω is the set of permissible control values), such that the equation

$$Q(x,T)=Q^*(x),\quad 0\le x\le S \tag{30}$$

is satisfied for a minimal time T.

However, in many cases eqn (30) cannot be accurately satisfied for any T. It then makes sense to formulate the problem as follows:

(*b*) To find a permissible control $U(t)$, $U\in\Omega$, $0\le t\le T$, where T is a fixed time, such that the functional

$$I=\int_0^S[Q^*(x)-Q(x,T)]^\gamma\,dx \tag{31}$$

which characterizes the measure of deviation of the actual distribution from the given one (γ a positive even integer), should reach a minimum.

Using the straight-line method, problems (*a*) and (*b*) may be reduced to an ordinary problem of optimum control for systems with lumped parameters.

In fact, splitting the interval [0, S] on the x axis into n equal parts by the points $x_0=0$, $x_1=s,\ldots,x_n=S$, where $s=S/n$, and replacing the second partial derivative of $Q(x,t)$ w.r.t. x in eqn (27) by the second difference ratio, we obtain a finite system of order $(n+1)$ of ordinary linear differential equations for the functions $q_i(t)$, $i=0,1,\ldots,n$:

$$\begin{aligned}\dot q_0&=-(\sigma+\beta)q_0+\sigma q_1+\beta U\\ \dot q_i&=\sigma(q_{i-1}-2q_i+q_{i+1}),\quad i=1,2,\ldots,n-1\\ \dot q_n&=\sigma(q_{n-1}-q_n)\end{aligned} \tag{32}$$

with the initial condition

$$q_i(0)=Q_0(is),\quad i=0,1,\ldots,n \tag{33}$$

and the final condition

$$q_i(T)=Q^*(is),\quad i=0,1,\ldots,n \tag{34}$$

Here β and σ are constant coefficients which can be expressed in terms of a and α.

In problem (*a*) the functional that has to be minimized is the time T. This problem can be solved by using the maximum principle. Gamkrelidze[3] has shown that its solution always exists and is unique.

In problem (*b*) the functional to be minimized is

$$I=\sum_{i=1}^{n}[q_i(T)-Q^*(is)]^\gamma \tag{35}$$

In certain cases it is required to determine the optimum variation law for a control action which is itself distributed in space, constraints being placed on it in time and also in spatial coordinates.

For example, it is sometimes material that too great spatial variations cannot be allowed in certain physical quantities such as temperature, pressure, electric field strength, etc.

We shall consider as an illustration the heat-exchange equation

$$b(y,t)\frac{\partial}{\partial t}Q(y,t)+b(y,t)v(t)\frac{\partial}{\partial y}Q(y,t)+Q(y,t)=U(y,t) \tag{36}$$

for the exchange between a stationary heating medium with temperature $U=U(y,t)$, $0\le y\le L$, $0\le t\le T$ (y being a spatial coordinate and t the time), and a material moving at velocity $v=v(t)\ge 0$ in the positive sense along the y axis and becoming heated in the process of moving over the interval $0\le y\le L$. The state of heating of the material is described by the function $Q(y,t)$. The initial and bondary conditions take the form

$$Q(y,0)=Q_0(y),\quad Q(0,t)=0 \tag{37}$$

In this case a permissible control is considered to be a function $U=U(y,t)$, $0\le y\le L$, $0\le t\le T$, that satisfies the conditions

$$A_1\le U(y,t)\le A_2 \tag{38}$$

$$A_3\le\frac{\partial}{\partial y}U(y,t)\le A_4 \tag{39}$$

where A_1, A_2, A_3 and A_4 are given constants.

Physically these constraints correspond to the fact that in feed-through heating installations one cannot allow too great amplitudes of temperature fluctuation in the heating medium, or excessive temperature drop over the length of the furnace.

In this case one has to determine the control $U=U(y,t)$, subject to conditions (38) and (39), such that, in spite of all possible disturbances of the heating process caused by variations in the velocity $v(t)$ and by variations in the thermal parameters $b(y,t)$ of the process, the deviation of the temperature of the material leaving the furnace from a certain given temperature Q^* should be on the average a minimum, i.e. one has to minimize the functional

$$I=\int_0^T[Q^*-Q(L,t)]^\gamma\,dt \tag{40}$$

where γ is a positive even integer.

In order to reduce the partial differential equations to difference differential equations, split up the interval [0, L] on the y axis into n equal parts by the points $y_0=0$, $y_1=1,\ldots,y_n=L$, where $l=L/n$. Replacing the partial derivative w.r.t. y in eqn (36) by the difference ratio, we obtain a system of order n of ordinary linear differential equations in the functions $q_i(t)$, $i=1,2,\ldots,n$,

$$b(il,t)\dot{q}_i+\frac{1}{l}b(il,t)v(t)[q_i-q_{i-1}]+q_i=U_i(t) \qquad (41)$$

with $q_0(t)=0$, $0\leq t\leq T$ and $q_i(0)=Q_0(il)$.

Equation (41) may be rewritten in the form

$$\dot{q}_i=\beta q_{i-1}+\alpha_i q_i+U_i, \quad i=1,2,\ldots,n \qquad (42)$$

where $U_i=U_i(t)=U(il,t)$, while the coefficients β and α_i can be expressed explicitly in terms of the functions $v(t)$ and $b(il,t)$.

According to conditions (38) and (39) the function $U_i(t)$, $0\leq t\leq T$, is subject to the constraints

$$A_1\leq U_i(t)\leq A_2, \quad i=1,2,\ldots,n \qquad (43)$$

$$A_3 l\leq U_{i+1}(t)-U_i(t)\leq A_4 l, \quad i=1,2,\ldots,n-1 \qquad (44)$$

The functional (40) must now be replaced by this one:

$$I=\int_0^T [Q^*-q_n(t)]^\gamma \,dt \qquad (45)$$

It can now already be seen that the maximum principle may be used for determining the optimum control actions $U_i(t)$, $i=1,2,\ldots,n$, $0\leq t\leq T$. In this case the permissible region Ω from which the values of the control vector $U(t)=U_1(t),\ldots,U_n(t)$ may be chosen is a closed convex polyhedron in n dimensional space, described by eqns (43) and (44).

Observe that the function $H(\psi_i, U_i)$ which has to be maximized according to the maximum principle in this case for each fixed t, $0\leq t\leq T$, takes the linear form in the U_i

$$H(\psi_i,U_i)=\sum_{i=1}^{n}\psi_i U_i \qquad (46)$$

Hence the problem of determining the optimum control actions $U_i(t)$, $i=1,2,\ldots,n$, at any instant t, $0\leq t\leq T$, reduces to the linear programming problem of maximizing the function H while satisfying conditions (43) and (44).

Thus it is that besides the accurate and quite general methods of solving optimal control problems which have a wide application in engineering, great significance is also attached to approximate methods of solving these problems, based on approximating partial differential equations by ordinary differential equations.

References

[1] BOLTYANSKIY, V. G., GAMKRELIDZE, R. V., and PONTRYAGIN, L. S. The theory of optimal processes. *Izv. Akad. Nauk SSSR, Seriya Matematicheskaya* 24, No. 1 (1960)

[2] BUTKOVSKY, A. G., and LERNER, A. YA. The optimal control of systems with distributed parameters. *Automat. Telemech.* 21, No. 6 (1960)

[3] GAMKRELIDZE, R. V. The theory of processes in linear systems that are optimal for rapid response. *Izv. Akad. Nauk SSSR, Seriya Matematicheskaya* 24 (1960)

[4] BUTKOVSKY, A. G. Optimal processes in systems with distributed parameters. *Automat. Telemech.* 22, No. 1 (1961)

[5] BUTKOVSKY, A. G. The maximum principle for optimal systems with distributed parameters. *Automat. Telemech.* 22, No. 10 (1961)

[6] BUTKOVSKY, A. G. Some approximate methods for solving optimal control problems on systems with distributed parameters. *Automat. Telemech.* 22, No. 12 (1961)

[7] PONTRYAGIN, L. S., BOLTYANSKIY, V. G., GAMKRELIDZE, R. V., MISHCHENKO, YE. F. *The Mathematical Theory of Optimal Processes.* 1961. Fizmatgiz

Programme Control and the Theory of Optimal Systems

YE. A. BARBASHIN

Introduction

Consideration is given to the system of differential equations

$$\frac{dx}{dt} = f(x, \eta, t) + u(c, y, t) \qquad (1)$$

where $x(t)$ is an n-dimensional vector, $y(t)$ an m-dimensional vector, $\eta(t)$ a certain (in general random) vector function, and c a constant vector. It is assumed that a certain trajectory $x = \psi(t)$ in pahse space is given for $0 \leq t \leq T$ $(0 < T \leq \infty)$. Assuming that certain information is received on the variation of $\eta(t)$, it is required to choose a vector c (problem A), or a vector function $y(t)$ (problem B), or a vector c and a function $y(t)$ (problem C), such that some solution of the system precisely or approximately realizes a motion along the trajectory $x\,\psi(t)$. The problem formulated in this manner is a problem of programme control.

Let O in *Figure 1* be the plant under control, whose object is to achieve a certain given mode of operation $x = \psi(t)$. To achieve this, a unit Y is introduced that develops a control u. In forming the control, use is made of information on the operating conditions $\psi(t)$ to be set up and also on external influences $\eta(t)$; this information may be received in distorted form for many reasons, e.g. delays and inertia in the transmission line C, measurement errors, random errors, etc.

If the problem had an accurate solution, then the required control would be determined by the relation

$$u(c, y(t), t) = \psi'(t) - f(\psi(t), \eta(t), t) \qquad (2)$$

Hawever, in a number of cases the system (2) cannot be solved for the control vector c or the control function $y(t)$. This may be due to the choice of an inadequate number of dimensions for the vectors c and $y(t)$, or the presence of incomplete or distorted information on the external influences $\eta(t)$. It may also happen that it is possible only to choose the control from some narrowly defined class of functions, such as piecewise-continuous functions, trigono-metrical polynomials, functions whose modulus has a constant limit, etc.

Thus the impossibility of solving system (2) accurately may lead to the statement of a number of problems of variational type. Bellman[1] considered such a problem when he used the dynamic programming technique to derive a control function $y(t)$ so as to make the maximum deviation of the system (1) trajectory from the required one a minimum.

One can state the problem of finding control functions that make a certain integral criterion of control quality a minimum. These problems may be solved by making use of the maximum principle of Pontryagin[2, 3], of classical variational principles[4], and of the principle of dynamic programming[5].

One can state the problem of minimizing the error with which the given trajectory satisfies system (1). If it is a question of a minimum mean-square error, then this statement of the problem leads to the simplest problems in the theory of mean-square approximations[6].

Finally, one can renounce all attempts to minimize the deviation, and seek a solution that simply gives a sufficiently accurate approximation. Thus, for example, Roytenberg[7] seeks a control function from the class of piecewise-continuous functions to give coincidence with the system (1) trajectory at a finite number of points on it.

It is observed that a distinction should be made between two essentially different cases in solving the problem of realizing the given process. In the first case the initial points of the actual and the required trajectories coincide; in the second the initial condition of the actual process may have any value. Normally in the second case the control is formed not only according to the magnitude of the disturbance but also according to the deviation of the controlled quantity from that required, i.e. in this case the control system will include feedback.

Programme Control and Optimal Systems

It should be noted that if the optimum principle is satisfied in one form or another in solving the problem of realizing a motion along a given trajectory, then in a number of cases it becomes possible to introduce feedback into the control system, which permits one to automatically correct the motion along that trajectory. This fact proves conclusively the advantage of systems designed on the basis of one or other optimum criterion. However, a number of difficulties are met in applying optimum criteria to the design of programme control systems and tracking systems. First of all, they often lead to designs requiring heavy computation or to designs that are technically difficult or impossible to execute. In this case, one must give up trying to satisfy the optimum principle, and restrict oneself to an approximate solution of the problem, with the aim of getting the best quality of fit within the bounds of technical possibility.

Usually in applying the optimum principle one needs a knowledge of the process to be realized over its whole duration. But it may happen that only certain statistical characteristics of the programmed process are known, or even that nothing is known about its future. In the latter case the optimal control theory is powerless, and the only reasonable approach is that of minimzing the deviation of the velocity vector for the current point from the tangent vector to a certain curve of pursuit from a given class, perhaps determined by a system of differential equations. Thus in this case the problem of minimizing the deviation of the given process from that required is replaced by the problem of minimizing the difference between two vector

fields, one of which determines the actual motion of the point and the other the required motion along a curve of pursuit. It should be noted that an analogous result is obtained if the control is chosen to give a maximum rate of decrease of a certain Liapunov function set up for the perturbed-motion system. The above approach, by introducing feedback, enables the deviation of the actual process to be rapidly reduced from that required, without the future course of the latter being known. It is shown below that an analogous result can be obtained by increasing the stability of the basic control circuit.

If the motion to be realized is known over the whole of its duration, then the following is the most natural method of solving the programme control problem. In the first stage a control that gives the most rapid means of reaching the given trajectory is sought, and in the second the control that achieves motion along that trajectory[8] is found.

The mathematical theory of optimal control that exists at present is basically a theory of optimal stabilization. This means that this theory permits, in the simplest cases, by the introductions of relay devices into the control system, an increase in the system's closed-circuit stability at zero input signal. In other words, the quality of the system's stability is heightened, using optimum criteria, irrespective of the nature and type of input actions that are processed. Clearly such a system will deal with input actions in various manners according to their structure. Below is given an example of a servo-system that reacts well enough to step-function inputs, and consequently also to any slowly varying inputs. However, in order to obtain this good qualiy tin the system, the requirement for optimum closed-circuit stability had to be abandoned.

Example

The control system having the block diagram shown in *Figure 2* is considered. Here f is the input signal, x the output signal, K the amplifier unit, A the unit forming the gain κ, which in general is variable. The problem is to find the optimum law of variation, given the constraint $|\kappa| \leq \kappa_0$, and the condition that the error ε decreases in some sense in the fastest way. Since in this case the time for complete elimination of the error is infinite from a mathematical point of view, the time for the error to fall within a given region surrounding the origin has to be discussed. Rapid action of this type will be called relative rapid action in distinction to the normal type.

The case where the plant under control is specified by an equation of the second order is considered, i.e. where $L(p) = p^2 + ap + b$. In this case the differential equation being examined takes the form

$$\ddot{\varepsilon}+a\dot{\varepsilon}+b\varepsilon=\ddot{f}+a\dot{f}+bf-\kappa\varepsilon \tag{3}$$

First consider the case where the external input f is absent, and find the law of variation of κ that gives relative rapid action. According to the results of Yemelyanov and Fedotova[9], the gain should be determined by the formula

$$\kappa=\kappa_0 \operatorname{sgn}\varepsilon(Tp+1)\varepsilon \tag{4}$$

where T^{-1} is the negative root of the equation

$$\lambda^2+a\lambda+b-\kappa_0=0 \tag{5}$$

Now consider the case where f is a step function, i.e. let $f=0$ for $t<0$ and $f=f_0$ for $t\geq 0$, assuming that the magnitude f_0 of the step cannot be measured. Reserving the freedom to choose T, take the previous switching law given by eqn (4). Clearly if $\kappa=\kappa_0$, eqn (3) after the step has passed will take the form

$$\ddot{\varepsilon}+a\dot{\varepsilon}+(b+\kappa_0)(\varepsilon-\varepsilon_2)=0 \tag{6}$$

where $\varepsilon_2 = bf_0/(b+\kappa_0)$.

Correspondingly, for $\kappa=-\kappa_0$ one gets

$$\ddot{\varepsilon}+a\dot{\varepsilon}+(b-\kappa_0)(\varepsilon-\varepsilon_1)=0 \tag{7}$$

where $\varepsilon_1 = bf_0/(b-\kappa_0)$.

Assuming that κ_0 is large enough a qualitative plot in the phase plane can be drawn for each of these equations without difficulty. Equation (6) in the phase plane corresponds to a family of spirals converging to a focus-type special point $(\varepsilon_2, 0)$. Equation (7) in the phase plane corresponds to a family of integral curves of hyperbolic type, with a 'saddle'-type special point $(\varepsilon_1, 0)$ through which pass two integral straight lines whose gradients are the roots of eqn (5).

Assuming now that the switching law is given by eqn (4), the phase diagram shown in *Figure 3* is obtained, provided only that $T\lambda_1 < -1$, where λ_1 is the negative root of eqn (5) is assumed. If the latter inequality is not satisfied, an obviously unsatisfactory result is arrived at, since the switching line (T) given by the equation $T\dot{\varepsilon}+\varepsilon=0$ will be cut by the integral curves over the whole of its length with $\varepsilon<0$, while in our case the straight line T is a sliding line everywhere except over the segment EF, where E and F are the points of contact with the integral curves corresponding to eqns (6) and (7). Thus the switching line resulting from the relevant optimum criteria will have been deliberately abandoned. If the representative point M falls to the left of the line T, then it will slide along this line as far as F, follow a curve of hyperbolic type as far as the line $\varepsilon=0$, then a spiral as far as the right-hand part of line T, where it will again start to slide towards E. On arriving at E it will approach the point $(\varepsilon_2, 0)$ along a spiral if $a>0$, while if $a<0$ it will start to move along a cycle consisting of the segment GE of line T and the segment EHG of the spiral. Thus any point in the plane arrives, eventually, either within a sufficiently small region about the point $(\varepsilon_2, 0)$ or at a limiting cycle corresponding to some self-oscillatory mode. It should be observed that the amplitude of the resulting self-oscillations is of the same order as $\varepsilon_2 = bf_0/(b+\kappa_0)$, and consequently can be made as small as required by increasing κ_0.

It should be noted that by increasing T the length of the segment over which it cuts the integral curves is decreased, but the speed of sliding along this line is also lessened since, as can readily be seen, the sliding law is given by the relation $\varepsilon=\varepsilon_0 \exp(-t/T)$. Thus proceeding from various quality criteria and combining speculation with experiment a reasonable value for the time constant T can be selected.

Toge ther with R. M. Yeydinov and I. N. Pechering the author has been carrying out analogous investigations for a third-order system. Here the main difficulty lies in the problem of synthesizing a corresponding optima lsystem.

Connection with the Accumulated Disturbance Problem

Returning now to the problem formulated in the first paragraph; as far as the approximation to the final section of

the trajectory is concerned, our problem is directly related to that of Bulgakov[10] on the accumulation of disturbances in a dynamic system.

Introducing the substitution $z = x - \psi(t)$ into the system of eqn (1), we transform it into the form

$$\frac{dz}{dt} = Z(z, \eta, t) + r(c, y, \psi(t), \eta(t), t) \tag{8}$$

where

$$Z(z, \eta, t) = f(z + \psi(t), \eta(t), t) - f(\psi(t), \eta(t), t)$$

$$r(c, y, \psi(t), \eta(t), t) = f(\psi(t), \eta(t), t) - \psi'(t) + u(c, y, t) = r(t)$$

System (8) is a system of equations for perturbed motion, the function $r(t)$ determines according to eqn (2) the approximation error of the programming or control functions, and the deviation of the solution $z(t)$ of system (8) from zero coincides with the deviation of the solution $x(t)$ of system (1) from the given function $\psi(t)$.

If system (8) is linear, then for $z(0) = 0$, $0 < t \leq T < \infty$ we have $z(t) = Ar(t)$, where A is a linear-bounded operator transforming the function $r(t)$ into the functions $z(t)$. If $\|A\|$ is the norm of the operator, then $\|z(t)\| \leq \|A\| \, \|r(t)\|$ is obtained. The latter relation is also the most general expression of the solution to the problem of disturbance accumulation. By taking various norms for $r(t)$ and $z(t)$ and computing $\|A\|$, the actual inequalities that solve this problem[11, 12] are obtained.

Connection with the Theory of Approximations

If as an optimum criterion that of the minimum error $r(t)$ (in any dimension) is taken, then the problem of realizing the given trajectory reduces to a problem in the theory of approximations. This problem is most effectively solved in the case where $r(t)$ depends linearly on the programming parameters and functions, and where we require a minimum of the mean-square approximation error. In this case the elementary rules of the theory of mean-square approximations are used for computing the control. It should be observed that here two essentially different cases are met. In the first case, by selecting the programming parameters from a sufficiently large number of them the approximation error can be made as small as required, i.e. realizing the given motion as accurately as necessary. In the second case the error of approximation cannot be made less than a certain value. Here it is worth while to state the problem of simultaneously choosing optimal values for the parameters and optimal programming functions. The success of such a choice depends, roughly speaking, on how well the given trajectory fits into linear subspaces in the various dimensions[11].

Trajectory Realization and Stability Theory

If it is wished approximately to realize motion along a given trajectory for the whole interval $0 < t < \infty$, certain difficulties arise. It can readily be seen that such an approximate realization is possible if the zero solution of the system

$$\frac{dz}{dt} = Z(z, \eta, t) \tag{9}$$

is stable in relation to continuously acting disturbances that are limited relative to the dimension in which the approximation error $r(t)$ is evaluated. There exist[13] stability criteria related to continuously acting disturbances limited in modulus or in mean value. Stability criteria can easily be deduced[14] for use with continuously acting disturbances limited in their mean square, which are of most interest in our problem. However, in solving the problem it was required to find convenient evaluations of continuously acting perturbations that were simultaneously evaluations of approximation errors.

Such evaluations[14] were found and it turned out to be best to make them in the dimension of space M with norm

$$\|r(t)\|^2 = \sup_{0 < t < \infty} \int_{kT}^{(k+1)T} |r(t)|^2 \, dt$$

where $|r(t)|$ denotes the length of the vector $r(t)$. Massera was the first[15] to point out the important role of the space M in stability theory.

Dwelling further on a question related to stability theory, the operating mode $\psi(t)$ is called stable in relation to the system $\dot{x} = X(x, t)$ if the zero solution of the system

$$\dot{z} = X(z + \psi(t), t) - X(\psi(t), t)$$

is asymptotically stable. From the preceding argument it is clear that only stable operating modes can claim to give a good approximation. Unfortunately few criteria for operating mode stability have so far been derived in relation to this system. Clearly if the basic system is linear and asymptotically stable, then any operating mode will be stable relative to it. The same property is possessed by the systems considered by Krasovskiy in his paper[16] (theorem 3.1). These systems are determined by the fact that for each of them a constant symmetrical matrix A can be defined having positive eigenvalues and such that the symmetrized matrix

$$[B_{ik}] = \left[\left(A\frac{\partial X}{\partial x}\right)_{ik} + \left(A\frac{\partial X}{\partial x}\right)_{ki}\right], \left(\frac{\partial X}{\partial x}\right)_{ik} = \frac{\partial X_i}{\partial x_k}$$

has negative eigenvalues μ_i satisfying the inequality $\mu_i < -d$, where $d > 0$ at all points of the space $-\infty < x_i < \infty, 0 \leq t < \infty$.

The interesting result obtained by Letov[17] is also noted, concerning non-linear control systems with parameters that vary only slightly. He has proved for a large class of systems of great importance in control engineering that the stability of a given operating mode implies the stability of all sufficiently close modes. In this case the closeness of the modes is assessed by the magnitude of the modulus of the difference between the programming functions.

Probably further results in this direction can be obtained on the basis of both existing and new criteria for asymptotic stability of linear systems with variable coefficients. It can easily be verified that in the unidimensional case Krasovskiy's criterion is a necessary and sufficient condition for the stability of any mode. It would be interesting to know to what extent this criterion is necessary for systems of a higher order.

Realization of Periodic Motions

Now let the right-hand side of the system (1) and also the function $\psi(t)$ be periodic in t with period T. Assuming that the

zero solution of system (9) is asymptotically stable to a first approximation, we can again formulate the conditions for a given motion to be realizable with the required accuracy. But in this case these conditions can be set more simply, since here the dimension in space M is given by

$$\|r(t)\|^2 = \int_0^T |r(t)|^2 \, dt$$

Furthermore it can be shown that even in the presence of an approximation error different from zero there exists an asymptotically stable periodic motion lying within an ε neighbourhood of the given periodic motion.

It should be observed that the results obtained can be extended without difficulty to the case where the motion to be realized is discontinuous, or more accurately has discontinuities of the first sort[14]. In this case the programming functions will appear as the sums of ordinary functions and linear combinations of δ functions.

Programme Control of Random Processes

Up to now attention has not been directed to the external influence or, more precisely, disturbance $\eta(t)$. Normally $\eta(t)$ is a random function, and so the actual mode of operation will be a random process. Naturally in this event the programmed mode also is random. The extension of the preceding results to the case of stochastic differential equations presents no difficulties, provided the following points are borne in mind. A random quantity, as is known, may be determined as a measurable function defined in some choice space $\overset{\circ}{\perp}$ (or space of elementary events). It is easy to see that the space $\overset{\circ}{\perp}$ can be constructed in such a way that it is the choice space for all random functions $\eta(t)$, $\xi(t)$ and $x(t)$ occurring in the equation

$$\frac{dx}{dt} = f(x, t, \eta(t)) + u(t, \xi(t)) \qquad (10)$$

where $\xi(t)$ is the distortion of the disturbance $\eta(t)$ (see *Figure 1*).

If a norm is defined by any means in the linear space of random quantities (as in the space of measurable functions defined in the choice space $\overset{\circ}{\perp}$), then differential eqn (10) is transformed into a differential equation given in the linear normalized space R, whose elements are random vectors. Here one should take as initial vectors in the solution of Cauchy's problem not only deterministic vectors but also any other random vectors from R, while the derivative and integral of a random function *w.r.t.* t should be understood as the derivative and integral in Bochner's sense. In particular, if as the square of the norm of a random vector the mathematical expectation of the square of the length of the vector is taken, then the concept of the derivative and integral of a random function coincides with the generally accepted one.

It should be observed that the theory of differential equations in a Banach space is well developed at the present day. By making use of this theory, one can readily formulate conditions for the existence, uniqueness and extensibility of solutions[18], and consider questions of stability[19] or questions of the existence and research of periodic motions[20]. All this enables the setting up of a completely analogous statement of the problem of realizing random processes and to obtain results identical to those presented above[21].

The reduction or elimination of the effect of disturbance by continuous tracking of it has found wide application in the theory of automatic control, mainly in the theory of composite control systems. This theory uses the so-called invariance principle developed by Academicians Luzin and Kulebakin, which has served as the starting point for a large number of papers on automatic control theory that have important applications.

Realization of Processes by means of Systems with Many-valued Characteristics

Barbashin and Alimov[22] have shown how to reduce systems of differential equations with relay-type hysteresis, and in general many-valued characteristics to a differential equation in a normalized linear space. Thus in this case also all the preceding results can be obtained by the same method as was indicated for the programming of random processes.

Conclusions

It has been seen in this paper that the accuracy of approximation to the trajectory depends on the degree of stability of the zero solution of the system (10). The better this stability is, as judged by any of the existing quality criteria, the smaller effect will approximation errors have on the deviation of the trajectory from the given one. Thus the problem of improving the response of programme control turns on the problem of increasing the stability of motion. Here, in particular, the theory of programme control again comes into contact with the theory of optimal control.

References

1 BELLMAN, R. Notes on control processes, Pt I. On the minimum of maximum deviation. *Quart. appl. Math.* 14 (1957)

2 PONTRYAGIN, L. S., BOLTNYANSKIY, V. G., GAMKRELIDZE, R. V., and MISHCHENKO, YE. F. *The Mathematical Theory of Optimal Processes.* Fizmatizdat (1961)

3 ROZONOER, L. I. Pontryagin's maximum principle in the theory of optimal systems, Pt II. *Automat. Telemech,* 20, No. 11 (1959)

4 LETOV, A. M. The analytical design of controllers, Pt I. *Automat. Telemech.* 21, No. 4 (1960)

5 LETOV, A. M. The analytical design of controllers, Pt II. *Automat. Telemech.* 22, No. 4 (1961)

6 BARBASHIN, YE. A. On the approximate realization of motion along a given trajectory. *Automat. Telemech.* 22, No. 6 (1961)

7 ROYTENBERG, YE. N. Some problems in the theory of dynamic programming. *Prikl. Matem. Mekh.* 23, No. 4 (1959)

8 BARBASHIN, YA. A. On a problem in the theory of dynamic programming. *Prikl. Matem. Mekh.* 24, No. 6 (1960)

9 YEMELYANOV, S. V., and FEDOTOVA, A. I. The design of optimal automatic control systems of the second order using limiting values of the elements of the control circuit. *Automat.Telemech.* 21, No. 12 (1960)

10 BULGAKOV, B. V. On the accumulation of perturbations in linear oscillatory systems with constant parameters. *Dokl. Ak. Nauk SSSR* 51, No. 5 (1946)

11 BARBASHIN, YE. A. The evaluation of the mean-square deviation from a given trajectory. *Automat. Telemech.* 21, No. 7 (1960)

12 BARBASHIN, YE. A. The evaluation of the maximum of the deviation from a given trajectory. *Automat. Telemech.* 21, No. 10 (1960

[13] GERMAIDZE, V. YE., and KRASOVSKIY, N. N. On stability in the presence of continuously-acting perturbations. *Prikl. Matem. Mekh*, 21, No. 6 (1957)

[14] BARBASHIN, YE. A. On the construction of periodic motions. *Prikl. Matem. Mekh.* 15, No. 2 (1961)

[15] MASSERA, J. L., and SCHÄFFER, J. J. Linear differential equations and functional analysis, Pt I. *Ann. Math.* 57, No. 3 (1958)

[16] KRASOVSKIY, N. N. Stability with large initial perturbations. *Prikl. Matem. Mekh.* 21, No. 3 (1957)

[17] LETOV, A. M. *The stability of non-linear controlled systems.* 1955. GITTL

[18] KRASNOSELSKIY, M. A., and KREIN, S. G. Non-local existence theorems and uniqueness theorems for systems of ordinary differential equations. *Dokl. Akad. Nauk SSSR* 102, No. 1 (1955)

[19] MASSERA, J. L. Contributions to stability theory. *Ann. Math.* 64, No. 1 (1956)

[20] MASSERA, J. L., and SCHÄFFER, J. J. Linear differential equations and functional analysis, Pt II. Equations with periodic coefficients. *Ann. Math.* 69, No. 1 (1959)

[21] BARBASHIN, YE. A. Programme control of systems with random parameters. *Prikl. Matem. Mekh.* 25, No. 5 (1961)

[22] BARBASHIN, YE. A., and ALIMOV, YU. I. Contribution to the theory of dynamic systems with non-single-valued and discontinuous characteristics. *Dokl. akad. Nauk SSSR* 140, No. 1 (1961)

[23] BARBASHIN, YE. A., and ALIMOV, YU. I. Contribution to the theory of relay-type differential equations. *Izv. Vyssh. Ucheb. Zav. Matemat.*, No. 1 (26), (1962)

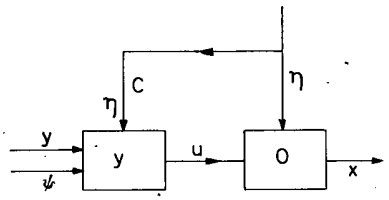


Figure 1

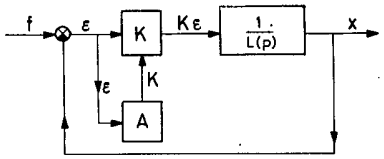


Figure 2

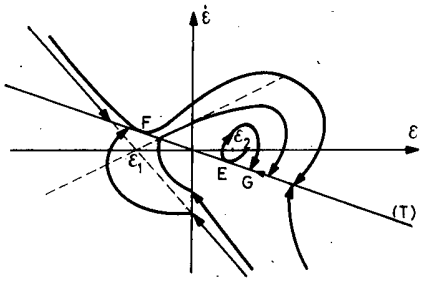


Figure 3

On the Searching of Extrema of Functions in Automatic Control Systems

A. A. VORONOV and M. B. IGNATJEV

This paper considers a distinctive approach to the problem of synthesis of local systems for automatic search of extrema of functions of many variables. The principle involved in construction of systems which react upon the partial derivatives of the sought function by the coordinates of the controlling devices is not new.

In the search of the extremum of the function of a single variable the problem is sufficiently definite; however, when the function depends upon several variables, the definiteness is lost and the solution of the problem becomes multi-valued. The function of a single variable $y = f(x)$ is represented by a plane curve, and if at a certain point x_1 one determines dy/dx, then it is necessary to vary x in order to approach the required extremum. A function of two variables may be represented by a surface $y = f(x_1, x_2)$. In this case, the path followed in passing from a given point to the point of extremum, while remaining on the surface, is not a singular one, but one of infinite multitude.

In 1959 Krasovskiy considered systems which searched the path to the extremum by the gradient method[7]. He also showed that depending upon the form of the surface $y = f(x_1, x_2, \ldots, x_n)$ the gradient method may be varied by making the shift of the controlling device x_i dependent on the derivative $f_i = \partial f/\partial x_i$ and also upon the derivatives of the function f with respect to other coordinates.

Approximately at the same time the Electro-Mechanical Institute of Leningrad considered the problem of simulating functions of many variables by means of digital differential analysers. This problem arose in connection with the construction of systems of programme control of metal cutting machines, first for simulating plane curves and then for curves lying on a given surface. The method which was utilized in this instance made it possible to indicate the general methods of synthesis with DDA (digital differential analyser) intended for simulating various forms of multidimensional surfaces, and also to indicate the quite general method of constructing systems of searching of extrema of functions of many variables, based on the principle of measurement of partial derivatives. The gradient method is obtained in this instance as a special case. This method also permits the searching of extrema, taking into account the boundaries at the intersection of multidimensional surfaces.

Before proceeding to the treatment of this method, it is necessary to consider the question of the structure of differential equations whose solution lies at the given intersection of multidimensional surfaces[4, 5].

The Structure of Differential Equations whose Solution Lies at the Intersection of Multidimensional Surfaces

The problem of finding differential equations whose solution is a given function is not a single-valued problem and its rational solution depends upon the means which are used for composing the sought system of equations, and upon which properties of the functions are utilized in the solution.

Indeed, in mathematical analysis there are given proofs of theorems on the singularity of solution of differential equations under given conditions; however, it is obvious that the converse problem has a multitude of solutions, that is, it is possible to find a multitude of differential equations whose singular solution will be the given function.

At the present time there exist two approaches for solving differential equations. The first approach permits the construction of an equation by introducing a parameter, and in the following this approach is called the parametric method of synthesis. The second approach of developing the method of synthesis is one which converts an equation into an identity, which equation is obtained by differentiating the output function with respect to the parameter.

Let there be a function

$$F(x_1, x_2, \ldots, x_n) = 0 \tag{1}$$

and an argument φ. In the parametric method of synthesis one finds first of all the parametric equations

$$x_i = x_i(\varphi), \qquad i = 1, 2, \ldots, n \tag{2}$$

which satisfy (1), and from these equations differential equations are found whose solution will be given by functions (2). In this case it is possible to find the differential equations by the method of $K(D)$ transform proposed by Kulebakin.

By the second method[4, 5] the parametric expression (2) is not sought, but the differential equations whose solutions satisfy (1) are immediately determined. As numerous observations indicate, the structures synthesized by the second method are considerably simpler than the structures synthesized by the first one.

The basis of this method of analytical construction of a differential analyser is the following lemma: having the function (1) of n variables which has a derivative in the given range of the variables, then in order that the solution of differential equations

$$\frac{dx_i}{d\varphi} = f_i, \qquad i = 1, 2, \ldots, n \tag{3}$$

under initial conditions satisfying (1) may transform eqn (1) into an identity, it is necessary and sufficient that the eqns (3) transform into an identity eqn (4):

$$\sum_{i=1}^{n} \frac{\partial F}{\partial x_i} \frac{dx_i}{d\varphi} = 0 \tag{4}$$

This lemma is then utilized for searching functions f_i: they are sought so that they may transform eqn (4) into an identity. This problem has a multitude of solutions, and this is what determines the fact that the problem of synthesis is not a single-valued one. No matter how we may determine f_i, they will in all cases be some functions of partial derivatives $\partial F/\partial x_i$. In the operation the functions f_i are sought out as linear functions of partial derivatives under the assumption that this is the simplest case.

As is shown[4, 5] the differential eqns (3) whose solutions satisfy (1) include arbitrary functions U_s whose number $s = C_n^2$, and the matrix of these arbitrary functions is symmetrical with respect to diagonal with zeros along the principal diagonal. For instance, the structure of differential equations whose trajectories are disposed on surface $F(x, y, z) = 0$ is determined by equations

$$\begin{aligned}\frac{dx}{d\varphi}&=u_1\frac{\partial F}{\partial y}-u_2\frac{\partial F}{\partial z}\\ \frac{dy}{d\varphi}&=u_1\frac{\partial F}{\partial x}+u_3\frac{\partial F}{\partial z}\\ \frac{dz}{d\varphi}&=u_2\frac{\partial F}{\partial x}-u_3\frac{\partial F}{\partial y}\end{aligned}\qquad(5)$$

where u_1, u_2, u_3 are arbitrary functions which determine the trajectory on the surface once they are given. They may be any functions as long as they satisfy Lipschitz conditions for right-hand sides of differential equations.

In simulating the trajectory at the intersection of surfaces

$$F_j(x_1, x_2, \ldots, x_n)=0 \qquad j=1,2,\ldots,m \qquad m<n \qquad (6)$$

the number of arbitrary functions u_s in the structure of differential equations is determined as

$$s=C_n^{m+1} \qquad (7)$$

and it is possible to determine the disposition of these arbitrary functions in the structure of the equations.

As an illustration of these methods of synthesis differential equations will be found whose solutions are disposed on surfaces

$$F_j(x_1, x_2, x_3, x_4)=0, \qquad j=1,2 \qquad (8)$$

At first the arbitrary functions are designated, the number of which in this case is $C_4^3 = 4$,

$$u_1=C_{123}, \quad u_2=C_{124}, \quad u_3=C_{134}, \quad u_4=C_{234}$$

The coefficient in which the subscript of the term C contains unity are disposed in the first line; the coefficients in which this subscript contains the number two, are situated in the second line, etc., that is,

$$\begin{aligned}\frac{dx_1}{d\varphi}&=u_1D_{23}^1+u_2D_{24}^1+u_3D_{34}^1\\ \frac{dx_2}{d\varphi}&=-u_1D_{13}^2-u_2D_{14}^2+u_4D_{34}^2\\ \frac{dx_3}{d\varphi}&=u_1D_{12}^3-u_3D_{14}^3-u_4D_{24}^3\\ \frac{dx_4}{d\varphi}&=u_2D_{12}^4+u_3D_{13}^4+u_4D_{23}^4\end{aligned}\qquad(9)$$

where the letter D designates the sum of the products of partial derivatives of the function (8) with respect to variables whose subscripts are present in the subscripts of the term D.

$$D_{12}=\frac{\partial F_1}{\partial x_1}\frac{\partial F_2}{\partial x_2}-\frac{\partial F_1}{\partial x_2}\frac{\partial F_2}{\partial x_1}, \qquad \text{etc.}$$

The superscript of the term D denotes the line.

The signs before the terms in eqns (9), as can be shown, are determined by the following rule: consider the order of the superscript and subscript of the symbol D, for instance that normally indicated by a pointer, and if there is an odd number of violations of the normal order, a minus sign is used before this term, and in other cases a plus sign is used. That is, for terms of the top line one has orders 123, 124, 134 in which there are no violations, and these are accompanied by a plus sign; in the second line there are 213 with one violation (2 being greater than 1), and a minus sign is used; 214 with a minus sign, 234—no violations—a plus sign; in the third line, 312—two violations—plus sign; 314—one violation—minus sign; 324—one violation—minus sign; in the fourth line, 412—two violations—plus sign, and 413 and 423 also have plus signs.

In an analogous manner one determines the structure of differential equations and the signs and dispostition of arbitrary functions in the latter by simulating trajectories at any number of intersecting surfaces with any number of variables.

It is of interest to note the presence of a maximum with respect to the number of arbitrary functions in the structure of differential equations for systems with a number of variables greater than six. The number of arbitrary functions for $n < 6$ decreases as the number of intersecting surfaces increases. For $n \geq 6$ the number of arbitrary functions for an increasing m at first increases, and only after having attained a maximum for $m = (n - 2)$ with even values of n, and for $m = (n - 2 \pm 1/2)$ for odd values of n, does it begin to decrease.

The arbitrary functions u_s in the structure of differential equations may be utilized as means of control in specifying prescribed motions on multidimensional surfaces, and as means of self-tuning of an automatic control system. Formula (7) relates the number of dimensions of the control space to the number of degrees of freedom of an automatic control system, and to the number of constraints imposed upon the system, while the presence of a maximum in the number of arbitrary controlling functions indicates an optimal structure as regards self-tuning of a holonomous system for $n \geq 6$.

Determination of Extrema of Functions

The problems of searching out the extrema of functions is one of the most widely encountered ones. There exist different methods of finding extrema in the presence of known partial derivatives, and different methods of automatic determination of these partial derivatives. However, at the present time, the methods of searching the extrema of functions in the presence of constraints placed upon the variables are not sufficiently well developed, and none of the existing methods assures that the motion to the extremum will proceed along a geodesic, or the shortest line.

In order to find the extrema one may utilize arbitrary coefficients in the structure of differential equations. Indeed, in order to assure the motion to an extremum–maximum with

respect to coordinate x_i, it is sufficient to prescribe such a motion that the coordinate x_i increases all the time, and this may be achieved by specifying the coefficients u_s in a proper manner. For instance, in order to attain a maximum with respect to z on the surface $F(x, y, z) = 0$ it is sufficient to assume in the system of eqns (5):

$$u_2 = a_1^2 \frac{\partial F}{\partial x}, \; u_3 = -a_2^2 \frac{\partial F}{\partial y}$$

In this case

$$\begin{aligned} \frac{dx}{d\varphi} &= u_1 \frac{\partial F}{\partial y} - a_1^2 \frac{\partial F}{\partial x} \frac{\partial F}{\partial z} \\ \frac{dy}{d\varphi} &= -u_1 \frac{\partial F}{\partial x} - a_2^2 \frac{\partial F}{\partial y} \frac{\partial F}{\partial z} \\ \frac{dz}{d\varphi} &= a_1^2 \left(\frac{\partial F}{\partial x}\right)^2 + a_2^2 \left(\frac{\partial F}{\partial y}\right)^2 \end{aligned} \qquad (10)$$

where $dz/d\varphi$ will be a positive definite form of a constant sign for all real values of x, y, z, which assures the stability of the process of finding the extremum in accordance with Liapunov[6]. At the point of the maximum with respect to z, the velocities with respect to all coordinates become zero. For a system of eqns (10) the point of maximum with respect to z proves to be a point of stable equilibrium. In the motion toward the extremum–minimum

$$u_2 = -a_1^2 \frac{\partial F}{\partial x}, \; u_3 = a_2^2 \frac{\partial F}{\partial y}$$

and $dz/d\varphi$ will be a negative definite form.

In an analogous manner we determine the coefficients u_s for a specified motion toward the extremum for surfaces with a large number of dimensions as well.

The synthesized structures may be utilized for searching out extrema of functions with any number of variables for individual surfaces as well as for cases in which constraints are taken into account, that is, for intersecting surfaces. For instance, in searching the maximum with respect to coordinate (x_4) at the intersection of surfaces (8) for a specified motion toward this extremum it is possible in the system of eqns (9) to let

$$u_2 = D_{12}, \; u_3 = D_{13}, \; u_4 = D_{23}$$

and $dx_4/d\varphi$ will be a positive definite form, and this fact assures stability of the process of searching the extremum according to Liapunov.

For a motion toward the extremum prescribed in this manner there remain free arbitrary functions in the synthesized structures, the number of which functions is equal to:

$$s = C_n^{m-1} - C_{n-1}^m$$

These free arbitrary functions may be utilized for simulating the trajectory during the time of the motion toward the extremum. In the example considered above there remains a free arbitrary function u_1 in the system of eqns (10). The free arbitrary functions may be utilized for prescribing the motion toward the extremum along a geodesic curve or one which is close to it.

It should be noted that all stationary points for the obtained differential equations will be points of equilibrium, but only points of the extrema will be points of stable equilibrium, while the saddle points will be points of unstable equilibrium.

If the number of intersecting manifolds is $m = n - 1$, then they determine a line in the n-dimensional space. In this case the problem is reduced to searching out the extremum in a one-dimensional manifold. It may be assumed that $d\varphi = \omega dt$, where t is the time and ω is an arbitrary function which satisfies the Lipschitz condition, and

$$\frac{dx_i}{dt} = \omega \xi_i(x_1, x_2, \ldots, x_n), \quad i = 1, 2, \ldots, n$$

For prescribing the motion toward the extremum in this case it is only necessary to specify the direction of the motion along the line. For example, in the motion toward the maximum with respect to x_n it is sufficient to assume that $\omega = \xi_n(x_1, x_2, \ldots, x_n)$, and then

$$\frac{dx_i}{dt} = \xi_n(x_1, x_2, \ldots, x_n) \cdot \xi_1(x_1, x_2, \ldots, x_n)$$

$$\frac{dx_n}{dt} = \xi_n^2(x_1, x_2, \ldots, x_n)$$

There follows a comparison of the described method of searching out the extrema and the gradient method. As shown by Krasovskiy[7], the gradient method assures stability, according to Liapunov, in the computing process of searching the extremum. This constitutes the similarity between them. But the gradient method assures the displacement toward the extremum only along some special trajectory, while the proposed method permits the variation of trajectory of motion toward the extremum.

Indeed, in the system of eqns (10) there remained one free arbitrary coefficient u which may be specified by a different method and which supplements the definition o ftrajectory for the motion toward the extremum. An analogous situation exists also in searching the extremum for other manifolds or their intersections, except for those which are one-dimensional. The gradient method constitutes a special case of the considered method of searching the extrema, when all the remaining arbitrary coefficients are set equal to zero; for instance, for the system of eqns (10), when $u_1 = 0$.

The remaining arbitrary coefficients may be prescribed in such a manner as to assure the motion toward the extremum along a trajectory which is optimal in some sense, including in this number a geodesic trajectory.

Figure 1 shows a block diagram of a system which searches out an extremum at the intersection of surfaces. The controlling signals produced by an analogue programming device (PD) are supplied to several simultaneously optimized plants 0_1, 0_2, ..., 0_n. On these plants the current values of partial derivatives which are supplied to the programming device are determined in some manner. The programming device constitutes a differential analyser (in particular, an electronic analogue installation) whose structure was described in the preceding paragraph. The setter of trajectories (ST) carries out such prescription of the arbitrary coefficients which remain free after the prescription of motion toward the extremum in order to assure the displacement toward it along some desired trajectory.

If the equations $F_i(x_1, x_2, \ldots, x_n) = 0$, $j = 1, 2, \ldots, m$ are known, then $0_1, 0_2, \ldots, 0_n$ are simply functional transforms. If only a part of these equations is known, this means that a part

of $0_1, 0_2, \ldots, 0_m$ are functional transforms (computer assemblies), while the other part are the plants.

In *Figure 2* is shown the block diagram of a system which utilizes the method of searching the extremum described above. As an example consider the case of searching the maximum on a surface $F(x_1, x_2, x_3, x_4) = 0$ with respect to coordinate x_4. The structure of the analogue device in this case is defined by equations

$$\begin{aligned}
\frac{dx_1}{d\varphi} &= u_1 \frac{\partial F}{\partial x_2} - u_2 \frac{\partial F}{\partial x_3} - \frac{\partial F}{\partial x_1}\frac{\partial F}{\partial x_y} \\
\frac{dx_2}{d\varphi} &= -u_1 \frac{\partial F}{\partial x_1} + u_4 \frac{\partial F}{\partial x_3} - \frac{\partial F}{\partial x_2}\frac{\partial F}{\partial x_y} \\
\frac{\partial x_3}{\partial \varphi} &= u_2 \frac{\partial F}{\partial x_1} - u_4 \frac{\partial F}{\partial x_2} - \frac{\partial F}{\partial x_3}\frac{\partial F}{\partial x_y} \\
\frac{dx_y}{\partial \varphi} &= \left(\frac{\partial F}{\partial x_1}\right)^2 + \left(\frac{\partial F}{\partial x_2}\right)^2 + \left(\frac{\partial F}{\partial x_3}\right)^2
\end{aligned} \tag{11}$$

In this instance we assume that $\partial F/\partial x_4 = -1$. The current values of partial derivatives may be determined by the method of synchronous detection. The considered system for $u_1 = u_2 = u_3 = 0$ is transformed into a scheme of extremal system cited by Krasovskii[7] and it differs from this scheme by the introduction of cross-links supplied to the input of the integrators. At the same time, the coefficients u_1, u_2, u_3 may be either constant magnitudes or functions of coordinates x_i, and be controlled by some index of the quality of operation of the system.

In specifying the motion along a geodesic curve in eqn (10), the free coefficient u_1, for instance, may be determined from the condition that for a geodesic curve the main normal to the curve coincides with the normal to the surface, and at the same time u_1 is determined as a complex function of coordinates.

If we search an extremum with respect to coordinate y on the surface $F(x, y, t) = 0$, where t is the time, then the structure of the analyser which specifies the motion toward the extremum will be defined by equations

$$\begin{aligned}
\frac{dx}{dt} &= -\frac{\partial F}{\partial x}\frac{\partial F}{\partial y} - u_2 \frac{\partial F}{\partial t} \\
\frac{dy}{dt} &= \left(\frac{\partial F}{\partial x}\right)^2 + \left(\frac{\partial F}{\partial t}\right)^2 \\
1 &= u_2 \frac{\partial F}{\partial x} - \frac{\partial F}{\partial t}\frac{\partial F}{\partial y}
\end{aligned}$$

As can be seen, by virtue of the last equation of this system of equations, the number of free arbitrary coefficients decreases.

It is possible to determine such constant coefficients u_s which assure the motion toward the extremum, perhaps not along the geodesic curve but at least along a path which is shorter than the trajectories followed during the motion toward the extremum by the gradient method, that is, when the free arbitrary coefficients are equal to zero. During the motion along a geodesic curve these coefficients in the general case will be complex functions. For constant free arbitrary coefficients, the technical realization of the proposed method is considerably simplified.

On the Search of Extrema of Functions in Automatic Control Systems

The operation involved in searching out extrema at the present time is automated to a large extent and may be used as a basis of construction of various automatic control systems. In the case of a limited range of change of variables the extremum may be sought taking into account the constraint.

$$\sum x_i^2 = R^2$$

The method described above permits this approach. Frequently in controlling chemical production of great complexity the problem of optimization of the free index of the quality of the process arises; for instance, if there is an object with a characteristic $F(x, y, z) = 0$, and it is required to determine such values of x, y, z which would provide an extremum to the free index $z' = l_3 x + m_3 y + n_3 z$ where l_3, m_3, n_3 are constant quantities, then this problem may also be solved on the basis of the method considered above.

Rewriting these equations using other designations, one has

$$F_1(x_1, x_2, x_3) = 0$$

$$F_2 = l_3 x_1 + m_3 x_2 + n_3 x_3 - x_4 = 0$$

The structure of differential equations whose solution lies at the intersection of these surfaces is determined as (9), where

$$D_{12} = m_3 \frac{\partial F_1}{\partial x_1} - l_3 \frac{\partial F_1}{\partial x_2}$$

$$D_{13} = n_3 \frac{\partial F_1}{\partial x_1} - l_3 \frac{\partial F_1}{\partial x_3}$$

$$D_{1y} = -\frac{\partial F_1}{\partial x_1}, \quad D_{23} = n_3 \frac{\partial F_1}{\partial x_2} - m_3 \frac{\partial F_1}{\partial x_3}$$

$$D_{2y} = -\frac{\partial F_1}{\partial x_2}, \quad D_{3y} = -\frac{\partial F_1}{\partial x_3}$$

The partial derivatives of the characteristic of the plant may be determined by some automatic method[1, 2].

The problem considered above may be formulated as a problem of searching an extremum in a given direction, which is characterized by coefficients l_3, m_3, n_3. At the present time an effort is being made to utilize the operation of searching extrema for solving the problem of constructing the motions[10]. The problem of constructing the motions based on energy levels[11, 12] may be formulated for the given kinematic scheme in terms of the intersections of the manifolds, and the motions themselves may be regarded as a solution of the problem of searching an extremum in a given direction.

In conclusion, consider the problem of possibilities of a global search. The finding of an extremal extremum requires a more thorough study of the investigated functions, and at the present time various strategies for solving this problem[1, 10, 13] have been proposed. One can propose yet another strategy for solving this problem as follows. Suppose that it is necessary to find the maximal maximum. Having investigated the function and having found several maxima, it is possible to pass a surface through them and the maximum of this surface will be

at least in the zone of gravity of the sought maximum of the maxima. If the approximated surface will have several maxima, then it may be smoothed in the same manner by finding the second approximated surface, etc. The number of approximated surfaces will be determined by the complexity of the investigated function.

References

[1] FELDBAUM, A. A. *Computers in Automatic Control Systems.* 1959. Moscow; Fizmatgiz

[2] IVANOV, V. N. On the determination of partial derivatives of functions of many variables in systems of automatic control. *Izvestiya AN SSSR, OTN, Energetika i avtomatika* No. 4 (1960)

[3] HOERL, A. E. *A Technique for Optimizing Process Conditions.* Fourth ASME/I.R.D. Conference, Newark, Delaware, April 1958

[4] IGNATEV, M. B. Synthesis of differential analysers for reproducing implicit functions. Collected works on problems of electromechanics. *Izd. AN SSSR* No. 5 (1961)

[5] IGNATEV, M. B. On the problem of synthesis of differential analysers. *Izvestiya AN SSSR, OTN, Energetika i avtomatika* No. 2 (1961)

[6] LIAPUNOV, A. M. *General Problem of Stability of Motion.* 1935 ONTI

[7] KRASOVSKII, A. A. Dynamics of continuous systems of extremal regulation based on the gradient method. *Izvestiya AN SSSR, OTN, Energetika i avtomatika* No. 3 (1959)

[8] IGNATEV, M. B. On the searching of extrema of functions with the aid of electronic analogs. *Reports of 4th Interuniversity Conference on the Application of Analogs in Various Branches of Technology.* Vol. 3, 1962. Izd. MEI

[9] IGNATEV, M. B. *Certain Problems of Synthesis and Application of Differential Analyzers as Control Devices.* Author's synopsis of dissertation. (1962)

[10] GELFOND, TZ. M., TZETLIN, M. L. On several methods of control of complex systems. *Progress of Mathematical Sciences (Uspekhi matematicheskikh nauk).* XVII, No. I (1962) 103

[11] GAMBARYAN, L. S. *Problems of Physiology of a Motion Analyzer.* 1962. Medgiz

[12] BERNSHTEYN, N. A. *On Construction of Motions.* 1947. Medgiz

[13] BOCHAROV, I. N., FELDBAUM, A. A. An automatic optimizer for searching the minimal of several minima. *Automat. telemek.* No. 3 (1962)

[14] VORONOV, A. A. *et al.* Digital analogs for systems of automatic control. *Izd. AN SSSR,* M. — L. (1960)

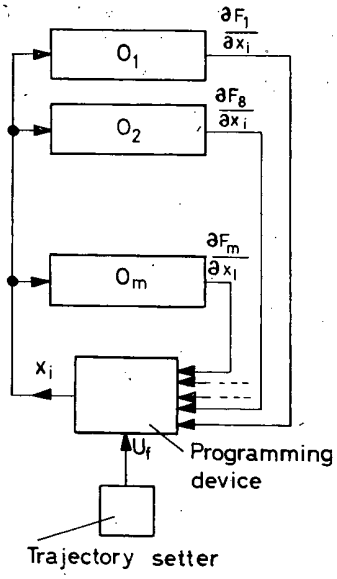


Figure 1

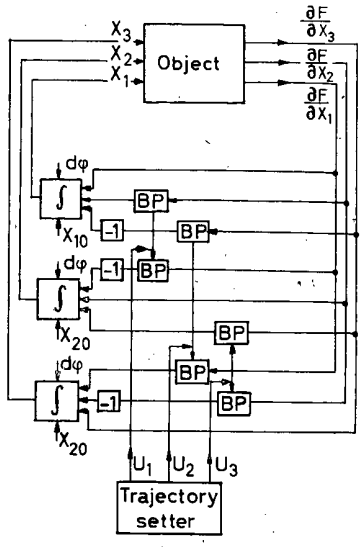


Figure 2

Automatic Systems with Learning Elements

G. K. KRUG and A. V. NETUSHIL

In the automation of continuous processes in the chemical industry (polymerization, fractional distillation, desiccation), in metallurgy (blast-furnace processing, rolling), in the paper, cement, food, and other industries, and also for the heat treatment of various materials, a number of difficulties are encountered which are due to incompleteness or lack of a mathematical description of the process.

Knowledge of the physicochemical laws determining the process gives only a qualitative idea of the principal relationships —insufficient for automatic control of the process.

Quantitative investigation of complex processes is carried out by experimental statistical methods[1–3]. As a result of mathematical treatment of sufficient information on the process, some equations of the connections between the parameters of a plant can be obtained. However, the presence of uncontrolled disturbances not only determines the probability character of these equations, but at the same time leads to the necessity of constantly examining them. In controlling such processes, the operator is guided to a considerable degree by past experience and intuition.

The algorithm of the functioning of automatic devices, designed for the optimal control of a process, must formally resemble in many respects the algorithm of control by a man. The operator, on taking control of the plant, has only the most general information on the nature of the actions required in various cases. As work proceeds, control experience is accumulated, and using past experiences definite tactics are worked out for use in the various situations. By constant improvement in methods of working, control of the plant is learned. In so doing, the results of experimental observations on the plant are often used, without going into the physical or chemical nature of the processes taking place.

With the object of obtaining fuller information on the controlled plant, the operator sometimes carries out test variations of the parameters according to a specific programme, and, having analysed the results, carries out the appropriate alteration to the method of operation.

Thus the control process can be presented in the form of a combined solution of two problems: (*a*) the study of the controlled pant, and (*b*) the control of the plant with the aim of obtaining the optimal behaviour.

Depending on the nature of the process, study of the plant can precede the development of a control algorithm, be periodically repeated during the control process, or be organically combined with the control process, providing continuous correction of the control algorithm.

Consider the set-up of the problem of control of the plant shown in *Figure 1*, classifying the parameters of the object of control as follows:

(1) *The set of primary controlled parameters of the process (the vector X)*—The magnitude of this set of parameters cannot be varied by the operator; for instance, the measurable characteristics of the input item or of incoming components, humidity, chemical composition, consistency, and certain indices characterizing the course of the process, such as change of the condition of the equipment, etc. Certain physical restrictions are laid upon the values of the primary parameters.

$$X \in \Gamma_X$$

where Γ_X is the region of possible values of X.

(2) *The set of secondary controlled parameters of the process (the vector K)*—This characterizes the state of the plant output, i.e., the quality of the end product (chemical composition, physical characteristics). There is a certain domain Γ_k where the values of K satisfy the prescribed requirements for the quality of the product.

(3) *The set of control parameters (the vector Z)*—The operator can influence this to vary the process, i.e. flow rate of water, fuel or raw material, conveyer speed, and pressure and temperature in different zones of the installation.

(4) *Index of the efficiency of the processes (I)*—In calculating I the cost price and productivity of the installation are taken into account. The productivity of the process alone can be taken as I.

(5) *The set of uncontrolled effects (the vector Y)*—This includes changes in characteristics of equipment owing to ageing, uncontrolled changes in quality of the input components, and uncontrolled variation of the process parameters, such as ageing of the catalyst.

In formal terms the process can be described as follows. At the plant input there occurs a variation of X. In the plant some alteration of the primary, control and uncontrolled parameters takes place, the results of which are felt at the plant output (K) after the expiry of the time of the technological cycle (τ_3) peculiar to the plant.

In the general case, the problem of controlling a process reduces to satisfying the following conditions:

$$K \in \Gamma_k \qquad I = I_{\max}$$

In discrete processes, for every cycle X and K, and also I, are constants. Each of the components of Z can be a time function with the interval $0 < \tau < \tau_3$. Minimizing the departure of this function from some prescribed mode often determines the quality of the product. The presence of uncontrolled factors and their nature are of the greatest importance in the choice of a control system.

The effect of uncontrolled factors can be partially reduced by the installation of a system of stabilizers of the various

process parameters, which neutralize the effect of random disturbances in the control network, and by the implementation of control by acting upon the corresponding settings of local stabilizers. Complete elimination of the effect of random factors is, however, theoretically impossible. Depending on the value of the uncontrolled factors, the control principles can be divided into the following three groups.

(1) Investigation according to a particular algorithm with the aim of establishing an optimal law of control, with subsequent realization of this law.

(2) Optimizing the process by means of a continuous, automatic predetermined search for the extremal regime.

(3) Optimization based on automatic statistical processing of experience of control, and application of the learning principle.

The first control principle finds application when it is possible to minimize the effect of the uncontrolled factors.

When the influence of the uncontrolled factors is considerable and the necessary information concerning the plant is lacking, the second or third principle of automatic optimization may be applied, according to the degree of complexity of the process.

Work is in progress in the laboratory of the Faculty of Automation and Telemechanics of the Moscow Institute of Power on the development of all three principles of self-adaptation of automatic control systems. Some of the questions the laboratory is working on are set out below.

Combined System of Programmed Control

It often happens that a definite relationship can be established between the quality of the product and some index of the regime. In these cases it is expedient to implement programmed control according to this index, thus securing the required quality of the product. This system is feasible if the index in question can be related to the number of controlled parameters X. When, owing to the complexity of the mathematical description of the plant, it is difficult to establish the required law of control according to the controlled parameter, the law must be found experimentally. One method for finding this law is to find a parameter Y, which is uncontrolled under normal control conditions, and is such that the quality of the product uniquely depends on it.

If, during the adjustment time of the process, it is possible to control the process temporarily by this index, and if there is a definite relation between this index and another controlled index by which it is possible to carry out control in normal operating conditions, then the solution of the problem can be found by a combined system of programmed control.

Let there be two indices of the course of the process, M and N. To obtain the requisite product quality it is sufficient that

$$N(\tau) \in N_H(\tau)$$

where $N_H(\tau)$ is the mode prescribed by technological considerations.

If during the adjustment time of the process it is possible to control the process by N, given a programme $N_H(\tau)$, it is possible to carry out a series of trials, storing the resulting law $M(\tau)$. After statistical treatment of a series of such functions $M(\tau)$, a law of control can be chosen by M and the required law specified $M_H(\tau)$.

Continuation of the process reduces to conventional programmed control by $M_H(\tau)$.

Thus in the first part of the control process

$$\begin{aligned} Z(\tau) &= N(\tau) \\ X(\tau) &= M(\tau) \end{aligned} \qquad (1)$$

In the second part of the process

$$Z(\tau) = M(\tau) \qquad (2)$$

This control principle is employed in a combined programmed controller for the process of induction tempering, in which $N(\tau)$ is the surface temperature of the item, measured by means of a thermocouple soldered on to it, and $M(\tau)$ is the voltage on the inductor. To obtain the required quality of tempering, heating must comply with a given law, for instance, rapid heating at constant speed followed by holding at constant temperature (curve 1 in *Figure 2*) or by slow heating at constant speed (curve 2 in *Figure 2*).

Exact calculation of the variation of the inductor voltage corresponding to the required temperature changes involves certain difficulties, since it is necessary to solve simultaneously three-dimensional Maxwell and Fourier equations for non-linear inhomogeneous media[4].

Setting the programmed temperature controller according to a specified law and 'remembering' the variation of inductor voltage in the process of temperature control (full line in *Figure 3*) make possible the determination of this law experimentally, and a programme can be drawn up for control of the tempering process by the inductor voltage (broken line in *Figure 3*). Programmed control of the inductor voltage dictates the course of the tempering temperature, thus ensuring the required quality.

The accuracy with which the temperature process is carried out depends on the extent of the influence of the uncontrolled factors (variation in material of the billets, of the current frequency, parameters of the generator, etc.). When these factors are relatively stable, this system for controlling the process provides the required quality of production.[5]

Control Systems Based on the Learning Principle

If the uncontrolled disturbances vary continually with time, automatic optimizers can be used, which carry out a predetermined search of the extremal regime[6–10].

The disadvantage of using optimizers is that often the algorithm realized by the system does not match the complexity of the problem of control of many processes simultaneously.

Consider a system of the learning type based on the principle of accumulation of positive control experience. It is assumed that from the dynamic point of view the plant is a non-linear element with pure time delay τ_3 for every pair of parameters affecting its input and output.

The block diagram of *Figure* 4 shows two interconnected blocks, one of which (Unit 1) supplies the control actions, in accordance with the method of operation of the process, in relation to the values of the primary parameters; that is, it realizes the principle of input control (by disturbance). To find the required law of control, feedback of the incentive type is introduced to signal the results of control, on the basis of the

values of the secondary parameters (incentive feedback are shown in the figures as a broken line).

Unit 2 takes into account the values of the secondary parameters, and trims Unit 1 in accordance with variation in characteristics of the plant. The functioning of Unit 2 is in some degree similar to that of automatic control systems which realize the principle of control by error. Application of the combined principle of control is the most promising method for the design of an automatic control device.

The information stored in the memories of Units 1 and 2 must, in an integral manner be a mathematical model of the process controlled in the best way in a predetermined manner. If the uncontrolled disturbances are of a varying nature, the mathematical model must constantly vary and adjust itself.

Two methods of information storage are possible—the table method and the formula method. With the former, the information is stored in the form of tables whose contents change in accordance with the algorithm governing writing and reading.

Table 1 shows a possible arrangement of the tables applicable to Unit 1 and Unit 2.

In the table for Unit 1 the values of Z are stored at the location of X. When coupled with the vector Z, the index I

Table 1. Information storage in automatic control device

Unit 1

X_1	$Z_1'I_1'$	$Z_1''I_1''$	
X_2	$Z_2'I_2'$	$Z_2''I_2''$	

Unit 2

ΔK_1	$\Delta Z_1'$	$\Delta Z_1''$	
ΔK_2	$\Delta Z_2'$	$\Delta Z_2''$	

defines the efficiency of the process for the prescribed combination of controlled primary parameters and control parameters.

In the table for Unit 2, the values of the vector ΔZ, that is, those values of variation of the control parameters which have restored the quality of the product from the state ΔK to the required level Γ_k, are written in the location of the vector ΔK which characterizes the departure of the quality vector of the end product from the specific value.

With the formula method, the quantitative connection between the controlled parameters is fixed in the form of a set of equations, in a polynomial form, for instance.

For Unit 1, the equation for one of the components of the vector has, in the general case, the form (for standardized values of the variables[2]):

$$z_i = a_1 X_1 + \ldots + a_m X_m + a_{m+1} X_1^2 + \ldots + a_{2m} X_m^2 + \ldots + a_N X_m^p \qquad (3)$$

or in more compact form

$$Z_1 = \sum_{j=1}^{N} a_j \prod_{k=1}^{m} X_k^{\alpha_{jk}} \qquad (4)$$

where

$$\sum_{k=1}^{m} \alpha_{jk} \leq p$$

For Unit 2 the equation for one of the components of the vector has the form:

$$\Delta Z_i = b_1 \Delta K_1 + \ldots + b_n \Delta K_n + \ldots + a_M \Delta K_n^S \qquad (5)$$

or in more compact form:

$$\Delta Z_i = \sum_{j=1}^{M} b_j \prod_{k=1}^{n} \Delta K_k^{\alpha_{jk}} \qquad (6)$$

where

$$\sum_{k=1}^{n_k} \alpha_{jk} \leq S$$

It may be presumed that, by means of a special algorithm for writing and reading the current information characterizing the controlled process, it can be said that the automatic devices incorporating either the formula or the table principle will be in some sense equivalent.

In whichever form the information may be stored, there must be an algorithm allowing processing of the incoming information in such a way that the contents of the memory express in the best way possible the current model of the controlled process. In the following an algorithm realizing the formula principle is considered.

Control Algorithm

The control algorithm is based on three coefficients: (*a*) prediction coefficient; (*b*) time weighting coefficient; and, (*c*) quality weighting coefficient.

Prediction Coefficient

The higher the power of the approximating polynomial, the more precisely can the main connections in the plant be described. However, with increase in the power of the polynomial there is a considerable increase both in the complexity of the programme and in the time for calculating the coefficients. A criterion is needed which makes it possible to evaluate quantitatively the precision with which the polynomial obtained approximates the actual relationship. This criterion is called the prediction coefficient since with its aid the dependability of the polynomial when used in the domain of the variable whichs have not yet been encountered can be evaluated.

The coefficients of the polynomial are determined from the minimum of the mean square of the approximation.

Use is made of the method of bringing a multiple non-linear correlation to the linear form[3].

The whole set of parameters X entering into each term of the polynomial eqn (4) is regarded as an independent parameter

$$\varepsilon_j = \prod_{k=1}^{m} X_k^{\alpha_{jk}} \qquad (7)$$

then eqn (4) has the form

$$Z_i = \sum_{j=1}^{N} a_j \varepsilon_j \qquad (8)$$

The auto-correlation coefficient of eqn (8) is found and expressed in terms of the auto- and cross-correlation coefficients of the variables

$$K_{ZZ}=\frac{\sum_{j=1}^{M} Z_j^2}{M}=\sum_{\substack{i=1\\j=1}}^{M} a_i a_j k_{\varepsilon_i\varepsilon_j} \tag{9}$$

where

$$k_{\varepsilon_i\varepsilon_j}=\frac{\sum_{i=1,\,j=1}^{M}\varepsilon_i\varepsilon_j}{M}$$

and M is the number of successive measurements of the variables. For convenience of analysis, eqn (9) is normalized, selecting as the norm the dispersion

$$D_{Z_0}=\frac{\sum_{j=1}^{M} Z_{0j}^2}{M} \tag{10}$$

where Z_0 is the observed value of the function.

Dividing eqn (9) by eqn (10) gives

$$\frac{k_{ZZ}}{D_{Z_0}}=\frac{\sum_{(i=1)\,j=1}^{M} a_i a_j k_{\varepsilon_i\varepsilon_j}}{D_{Z_0}} \tag{11}$$

If $Z_i = Z_{i0}$, both sides of eqn (11) are identically equal to zero.

In fact, the following inequality holds.

$$\theta=\frac{\sum_{i=1,\,j=1}^{M} a_i a_j k_{\varepsilon_i\varepsilon_j}}{D_{Z_0}}\leq 1 \tag{12}$$

since by determining the coefficients of the approximating polynomial by the method of least squares a 'smoothing' is produced, the magnitude of which depends on the power of the polynomial.

The quantity θ quantitatively expresses the degree of the probability prediction, i.e., the quality of the approximation.

Fixing a definite degree of prediction $(0-1)$ and passing successively from $p = 1$ to $p = 2{\cdot}3$ etc., the system will cyclically check the actual degree of prediction by eqn (12), and seek the correlation $\theta \geq \theta_0$.

It is also expedient to estimate the weight of each term of the polynomial. This can be done with the aid of the coefficient.

$$\beta_j=\frac{a_j k_{\varepsilon_j\varepsilon_i}}{D_{Z_0}}\ll 1 \tag{13}$$

Setting the minimal level β_0 for the coefficient β_j, it is arranged that after each operational cycle, β_i is calculated for all the terms. Terms with $\beta_i < \beta_0$ are eliminated from eqn (8), freeing the equation from weakly expressed connections of secondary importance.

The coefficients of the approximating polynomial are functions of the auto- and cross-correlation coefficients for the variables ε_i and the approximated quantity Z_{0i}[3].

Coefficient of Time Weighting

Owing to the unstable nature of the vector Y, the coefficients of the approximating polynomial must continually vary. The object of time weighting is to calculate the new values of the paired products of the variables which determine the correlation coefficient with a greater weight than the previous one, and gradually to forget the past values. The simplest method of time weighting is that of the sliding interval. With this method the correlation coefficient is calculated with respect to M previous values of the paired products

$$k_{\varepsilon_l\varepsilon_K}=\frac{\sum_{j=1}^{M}(\varepsilon_l\varepsilon_k)_{N-j}}{M} \tag{14}$$

where N is the serial number of the measurement.

The disadvantages of this method are the presence of a 'transient process' in the calculation of the correlation coefficient (for $N < M$), and the necessity for storing in the memory all the values of the paired products used for calculation.

A method of continuous weighting is possible, for which each paired product is multiplied by a weight function of the form

$$G_b=\alpha^{N-i} \tag{15}$$

where N is the serial number of the last cycle; i is the serial number of the information for which the weighting coefficient is being calculated; and $\alpha < 1$ is the coefficient of time weighting.

It can be shown that when the weighting function is introduced like this it is sufficient to store only the resulting value of the correlation coefficient for the $(N-1)$th cycle.

In fact:

$$(K_{\varepsilon_l\varepsilon_K})_N=\frac{\sum_{i=1}^{N}(\varepsilon_l\varepsilon_k)_i\,\alpha^{N-i}}{Nf(G_b)} \tag{16}$$

where $f(G_b)$ is a coefficient taking into account the attenuation of the information summed up in the numerator of eqn (16). Putting $(\varepsilon_l\,\varepsilon_k)_1 + (\varepsilon_l\,\varepsilon_k)_2 = \ldots = (\varepsilon_l\,\varepsilon_k)_N = \varepsilon_l\,\varepsilon_k$, the values of the coefficient

$$f(G_b)=\frac{\sum_{i=1}^{N}\alpha^{N-i}}{N} \tag{17}$$

Hence, substituting eqn (17) in eqn (16) gives

$$(K_{\varepsilon_l\varepsilon_k})_N=\frac{\sum_{i=1}^{N}(\varepsilon_l\varepsilon_k)_i\,\alpha^{N-i}}{\sum_{i=1}^{N}\alpha^{N-i}}=\frac{\alpha\sum_{i=1}^{N-1}(\varepsilon_l\varepsilon_k)_i\,\alpha^{N-i-1}+(\varepsilon_l\varepsilon_k)_N}{\sum_{i=1}^{N}\alpha^{N-i}} \tag{18}$$

Call the total coefficient of attenuation per cycle

$$L_N=\alpha^{N-1}$$

Then

$$L_N=\alpha\sum_{i=1}^{N-1}\alpha^{N-1}+1=\alpha L_{N-1}+1 \tag{19}$$

Bearing in mind that

$$(K_{\varepsilon_l\varepsilon_k})_{N-1}=\frac{\sum_{i=1}^{N-1}(\varepsilon_l\varepsilon_k)_i\,\alpha^{N-i-1}}{L_{N-1}}$$

write

$$(K_{\varepsilon_l\varepsilon_k})_N=\frac{\alpha L_{N-1}(K_{\varepsilon_l\varepsilon_k})_{N-1}+(\varepsilon_l\varepsilon_k)_N}{\alpha L_{N-1}+1} \qquad (20)$$

Thus to calculate the correlation coefficient in the Nth cycle the value of the correlation coefficient in the $(N-1)$th cycle must be stored, and also the total coefficient of attenuation in this cycle.

If N tends to infinity, the limit value $(K\varepsilon_l\,\varepsilon_k)$ is given by

$$(K_{\varepsilon_l\varepsilon_k})_N=\alpha(K_{\varepsilon_l\varepsilon_k})_{N-1}+(1-\alpha)\,\varepsilon_l\varepsilon_k \qquad (21)$$

since

$$\lim_{N\to\infty} L_{N-1}=\frac{1}{1-\alpha}$$

It must be remembered that the calculated correlation coefficients are modified indices of the interconnection of two random functions. It is therefore more accurate to call these coefficients pseudocorrelation coefficients[13].

Coefficient of Quality Weighting

Besides being weighted for time, the information arriving from the plant must be weighted for quality. In other words, evaluation of the information must depend on the magnitude of the technical and economic index to which it corresponds.

The introduction of quality weighting allows purposeful accumulation of information with deliberate 'reinforcement'. Although the polynomial so calculated approximates the interconnection between Z and ε in a distorted form, its value lies in the fact that it expresses the control problem.

In the general case, the quality index p is a function

$$p=\varphi(k,\prod)$$

Introducing the coefficient of quality weightung g, which depends on the value of the quality index p, $g=G_k(p)$

$$g=0 \quad \text{when} \quad p<p_0$$

$$1<g\leq 1 \quad \text{when} \quad p\geq p_0$$

where p_0 is some specified level of the quality index.

Taking into account weighting with respect to time and quality, the correlation coefficient is written in the form

$$(k_{\varepsilon_l\varepsilon_k})_N=\frac{\alpha\sum_{i=1}^{N-1}(\varepsilon_l\varepsilon_k)\,\alpha^{N-i-1}\,g_i+(\varepsilon_l\varepsilon_k)_N\,g_N}{\sum_{i=1}^{N}\alpha^{N-1}\,g_i} \qquad (22)$$

Using the notation

$$\sum_{i=1}^{N}\alpha^{N-i}\,g_i=R_N$$

as with eqn (20) gives

$$(k_{\varepsilon_l\varepsilon_k})_N=\frac{\alpha R_{N-1}(k_{\varepsilon_l\varepsilon_k})_{N-1}+(\varepsilon_l\varepsilon_k)_N\,g_N}{\alpha R_{N-1}+g_N} \qquad (23)$$

Selection of the Control Coefficients

The introduction of the coefficients θ_0, β_0, α, g makes it possible to work out the current mathematical description securing the best control of the process. The success of the work depends to a considerable extent on the correct choice of numerical values of the coefficients. The system must be capable of automatically varying the values of these coefficients in accordance with the variation of the statistical characteristics of the vector of the uncontrolled factors.

In fact, in the periods of time in which a variation of the vector Y takes place it is necessary to secure the quickest possible renewal of the memory (to decrease α), to reduce the power of the polynomial and simplify its form (to decrease θ_0 and β_0) and to increase the significance of a successful control trial (to increase g_0).

It appears that the difficulties of the problem are insuperable, since the vector itself does not enter into the polynomial in explicit form; however, some approaches to its solution may be noted.

As an indirect measurement of the variability of the vector, the mean square of the variation of all the correlation coefficients in one cycle of calculation can be taken

$$A^2=\frac{\sum_{i=1}^{\theta}[(k_{\varepsilon_l\varepsilon_k})_N-(k_{\varepsilon_l\varepsilon_k})_{N-1}]_i^2}{Q} \qquad (24)$$

where Q is the number of correlation coefficients subject to calculation, and

$$Q=\frac{m(m-1)}{2} \qquad (25)$$

where m is the number of terms of the approximating polynomial.

To average this evaluation and to eliminate the effect of disurbances of short duration, a quantity A, calculated from a finite number of cycles in the sliding interval of the average, can be used.

The evaluation can also be continuously averaged with the aid of time weighting, as in eqns (20) and (21).

Using the evaluation of the non-stationary vector A, it can be connected with (in linear form, for instance) the coefficients $\theta_0, \beta_0, \alpha, g$ which control the mathematical model of the controlled process.

Realization of the Algorithm of Control

As an illustration, *Figure 5* shows a programme of the functioning of a learning system of the formula type (without automatic trimming of $\theta_0, \beta_0, \alpha, g$).

The programme is realized by means of a digital computer. The programme envisages random search for ΔZ in the case when application of the recommendations held in Unit 2 does not lead to the required result.

The memory device is divided into two blocks, independently dealing with only the primary or only the secondary parameters of the process, in contemplation of linear introduction of the control parameters into the general connection equation;

that is, automatic devices of this design are best used with plants whose connection equations have the form

$$k_i = \sum_j [Z_j F_{1i}(X, Y) + F_{2i}(X, Y)] \qquad (26)$$

where F_{1i} and F_{2i} are any functions of the primary parameters and uncontrolled factors Y.

Depending on the characteristics of the process, and in particular on the nature of the variation of the uncontrolled factors (vector Y), and also depending on technical and economic considerations, the mode of operation may be continuous, cyclic or one-time.

With continuous operation, the automatic device is connected with the process permanently, and learning is continuous. This mode of operation is suitable for processes in which the uncontrolled factors vary continuously and substantially.

The cycle mode of operation consists of periodic connection of the automatic device to the plant for correction of the law of control. The interval of time during which automatic control is carried out by a rigid programme worked out by the automatic device is determined by the periodicity of the variation of the uncontrolled factors. In this method of operation, the device can serve several processes at the same time.

The one-time mode of operation can be successfully applied during running-in tests of technological processes for which variation of the uncontrolled parameters is small. A final mathematical description of the process worked out by the device is used for further control of the process in the form of flow charts, for the simplest programmed control systems realizing the law of control obtained.

References

1 VORODYUK, V. P., and KRUG, G. K. Finding connection equations in complex plants. *Automat. Telemech., Moscow* 11 (1961)

2 LUKOMSKII, YA. I. Correlation theory and its application to production analysis. [illegible]

3 HAL'D, A. Mathematical statistics with technical applications. [illegible]

4 NETUSHIL, A. V. The object of induction or radiation heating as an element in a control system. *Izv. OTN Energet. Automat.* 2 (1962)

5 KOLOMEITSEVA, M. B. A combined programmed control for an induction heating process. *Izv. OTN Energet. Automat.* 1 (196[illegible])

6 FEL'DBAUM, A. A. *Computers in Automatic Systems.* 1959. Moscow; Fizmatgiz

7 KRASOVSKII, A. A. The principles of search and dynamics of continuous optimalizing control systems. *Symp. on Automatic Control and Computer Technique* No. 4. 1961. Moscow; Mashgiz

8 KAZAKEVICH, V. V. Investigation of non-linear processes in an optimal regulator. Theory and application of discrete automatic systems. *Izd. AN SSSR* (1960)

9 NETUSHIL, A. V. Self-oscillation in discrete automatic systems. *Automat. Telemech., Moscow* 3 (1962)

10 UIDROU, B. A self-adapting sampled-data system. *Automatic and Remote Control.* 1961. London; Butterworths

11 NETUSHIL, A. V., KRUG, G. K., and LETSKII, E. K. Use of 'learning' systems for the automation of complex production processes. *Izv., VUZOV SSSR, Mashinostroyenie* 12 (1961)

12 KRUG, G. K., and LETSKII, E. K. A learning automatic device of the table type. *Automat. Telemech., Moscow* 10 (1961)

13 KALMAN, R. E. Design of a self-optimizing control system. *Amer. Soc. mech. Engrs* Pap. 57, RD 12 (1957)

14 IVANOV, A. Z., KRUG, G. K., KUSHELEV, YU. N., LETSKII, E. K., and SVECHINSKII, V. B. Learning-type control systems. *Proc. Moscow Power Inst.* 44

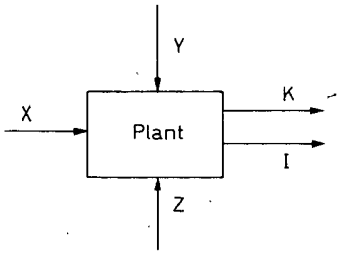


Figure 1. Plant

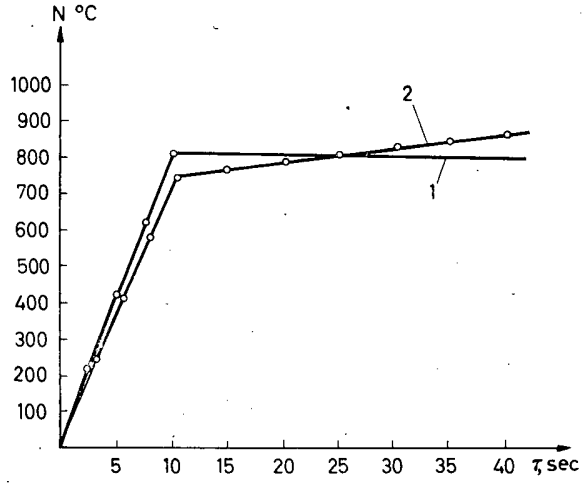


Figure 2. Heating characteristics of sample during tempering

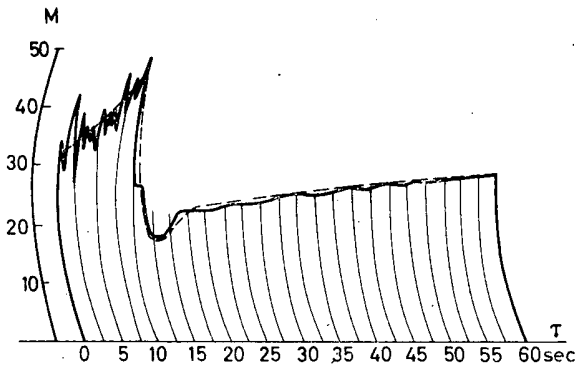


Figure 3. Variation of voltage on inductor

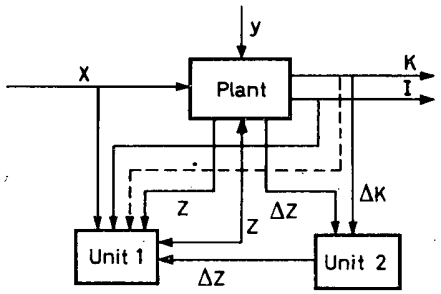


Figure 4. Block diagram of automatic device

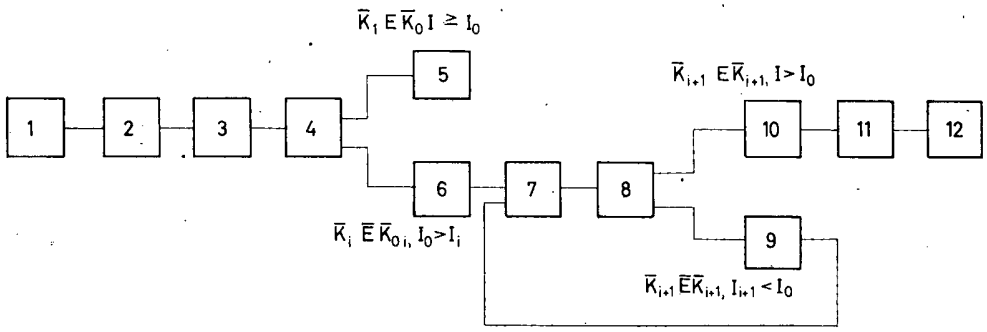


Figure 5. Programme of formula-type automatic device.

1: Input of parameters X_i; 2: calculation of parameters Z_i from eqn (3); 3: action upon plant; 4: input and analysis of K_i and I_i; 5: stop; 6: calculation of correction ΔZ_i from eqn (5); 7: action upon plant; 8: input and analysis of K_{i+1} and I_{i+1}; 9: random search for process input ΔZ_{i+1}; 10: conversion of $K_{\varepsilon_i \varepsilon_j}$ and D_{Z_0}; 11: change of control coefficient [eqns (3) and (5)]; 12: stop

Synthesis of Systems with the Fixed Characteristics of Equivalent Self-adjusting Systems

M. V. MEEROV

The problem of 'self-adjustment' in a control system arises in connection with the fact that in the operational process the characteristics of the control object may vary within a wide range. Under those conditions, the adjustment of the control system, or even its structure, which was entirely expedient for the initial form of the object's characteristics, may prove to be completely unsatisfactory for the altered characteristics.

A change in the characteristics of the object to be regulated may be conditioned, basically, on the two most prevalent factors. In the first place, a change in the characteristics may be brought about by external disturbances that are applied to the object; and in the second place, a change in the object's characteristics may take place in the course of its operation.

The problem of 'self-adjustment' also comes up in a number of cases where the information regarding the characteristics and properties of the object is insufficient; it is only known that the object's characteristics have an extremum for some qualitative criterion, and the control system's problem consists in a search for this extremum and in maintaining the object's operational conditions on this extremum. A rather large number of studies (for example, by Fel'dbaum[1, 2, 4], and by Doganovskii and Fel'dbaum[3]) have been devoted to methods of searching for, and adjustment of the system to, the disclosed extremum. In the present paper, methods are considered for the purpose of constructing systems with fixed characteristics that would maintain the most favourable adjustment, independently of external disturbances, and the character of which may be practically arbitrary. The sole limitation is the one regarding disturbances in accordance with the modulus (absolute value). In the present study, no consideration is given to the method of searching for the extremal condition for some qualitative criterion. If, however, the extremum is established by some method or other, then the structures examined below maintain this extremum automatically, without the need for a duplicate search.

Methods of Constructing Control Systems for the Case where the External Disturbances may be Measured

Consider the automatic control system whose schematic diagram is shown in *Figure 1*. In this diagram the designation $w_2(p)$ is given to the transfer function of the regulating object, $kw_1(p)$ and $w_3(p)$ to the transfer functions of the control system and of the stabilizing device, and F to the external disturbance. kw_1 and $w_3(p)$ have been selected in such a way that, in the absence of disturbances, the $F(p)$ process, which is the most favourable from the point of view of the selected qualitative criterion, is attained in the case of a sufficiently large amplification factor, k. Thus, for example, the optimum operating conditions are realized where there is an unlimited increase in the amplification factor, the object is non-linear, and there is a non-linear return communication in the optimum control circuit with an automatic potentiometer[5, 6]. It is natural for the designed circuit to remain stable where there is an unlimited increase in the amplification factor. The following situation is demonstrated: the structure, which is shown in *Figure 1*, upon giving no consideration to external disturbances and where k tends to infinity, is equivalent to the system in *Figure 2*, where consideration is given to disturbances and where k tends to infinity. In other words, in order to eliminate the effect of external disturbances that are capable of being measured, they should be supplied to the input of the stabilizing device in the form of an additional signal. Actually, the transfer function of the system in *Figure 1*, without calculating the external disturbances, will have the form:

$$k(p)=\frac{x_{\text{output}}(p)}{x_{\text{input}}(p)}$$

$$=\frac{\dfrac{kw_1(p)}{1+kw_1(p)w_3(p)}w_2(p)}{1+\dfrac{kw_1(p)}{1+kw_1(p)w_3(p)}}\cdot\frac{kw_1(p)w_2(p)}{1+kw_1(p)w_3(p)+kw_1(p)w_2(p)} \quad (1)$$

Assume that k tends to infinity; then,

$$k_{\text{adjusted}}(p)=\frac{w_1(p)w_2(p)}{w_1(p)w_3(p)+w_1(p)w_2(p)}=\frac{w_2(p)}{w_3(p)+w_2(p)} \quad (2)$$

Now, the transfer function for the circuit in *Figure 2* is found; one has:

$$Y(p)=kw_1(p)\{x_{\text{input}}-x_{\text{output}}-[Y(p)+F(p)]w_3(p)\} \quad (3)$$

$$x_{\text{output}}(p)=w_2(p)[Y(p)+F(p)] \quad (4)$$

From (3),

$$y(p)=$$

$$\frac{kw_1(p)x_{\text{input}}(p)-kw_1(p)x_{\text{output}}(p)-kw_1(p)w_3(p)F(p)}{1+kw_1(p)w_3(p)} \quad (5)$$

Substituting the value $y(p)$ from (5) and (4), one obtains either:

$$[1+kw_1(p)w_3(p)]x_{\text{output}}(p)$$

$$=kw_1(p)w_2(p)x_{\text{input}}(p)-kw_1(p)w_2(p)x_{\text{output}}(p)$$

$$-kw_1(p)w_2(p)w_3(p)F+w_2(p)F+kw_1(p)w_3(p)F$$

from which:

$$x_{output}(p)=\frac{kw_1(p)\,w_2(p)\,x_{input}(p)+w_2(p)\,F(p)}{1+kw_1(p)\,w_2(p)+kw_1(p)\,w_3(p)} \quad (6)$$

Where k tends to infinity, one obtains:

$$x_{output}(p)=\frac{w_1(p)\,w_2(p)\,x_{input}(p)}{w_1(p)\,w_2(p)+w_1(p)\,w_3(p)}$$

or

$$\lim\frac{x_{output}(p)}{x_{input}(p)}=\frac{w_2(p)}{w_2(p)+w_3(p)} \quad (7)$$

i.e. exactly the same expression as eqn (2). From what has been obtained it follows that the system in *Figure 2*, where there is a sufficiently large amplification factor, will behave as a self-adjusting one, in the sense that its characteristics will remain unchanged despite the presence of external disturbances whose character is practically unlimited.

Methods of Plotting Structures for the Case where it Does not Appear Possible to Measure Disturbances Directly

Now consider the case where the object's characteristics change due to the effect of external disturbances, but where it does not appear possible to measure these disturbances. Such a situation is, for all practical purposes, highly prevalent. A series of disturbances is difficult to measure, in the first place, because of the properties of the disturbances themselves, and in the second place because of the absence of sufficiently high-speed measuring devices for the measurement of the external disturbances.

The solution of the problem in the given case is carried out in the following fashion. Assume that the object's characteristics are known for the case where disturbances are absent. For this case, a control system is constructed in such a way that the optimum operational conditions should be attained when there is an unlimited increase in the amplification factor, k. Strictly speaking, in the absence of interferences, the system has the form shown by *Figure 1*. As was indicated earlier in this paper, in the case where k tends to infinity, one has:

$$k_{adj.}(p)=[w_2(p)]/[w_3(p)+w_2(p)]$$

Now *Figure 3* is plotted. The output of the controlling part of the circuit, which is designated in *Figure 3* by the letter y, is fed to the input of the real object and to the input of the model with the transfer function $w_2'(p)$. In future, $w_2'(p)$ is called the transfer function of an ideal object.

The difference between the outputs of the ideal and real objects is fed through a converting device with a transfer function $w_{convert}(p)$, to the input of the stabilizing device. Now the transfer function of the system in *Figure 3* is found.

$$y(p)=kw_1(p)\{x_{input}(p)-x_{output}(p)$$
$$-w_3(p)[Y(p)+(x_{output}(p)-x_{output}(p))\,w_{convert}(p)]\} \quad (8)$$

$$x_{output}(p)=w_2(p)[Y(p)+F(p)] \quad (9)$$

$$x'_{rect.}(p)=w_2(p)\,y(p) \quad (10)$$

On the basis of (9) and (10), one may write:

$$x_{output}(p)-x_{output}(p)$$
$$=w_2(p)[y(p)+F(p)-Y(p)-Y(p)]=w_2(p)\,F(p) \quad (11)$$

in calculating (11), eqn (8) is written as:

$$Y(p)=kw_1(p)\{x_{input}(p)-x_{output}(p)$$
$$-w_3(p)\,y(p)-w_3(p)\,w_{convert}(p)\,w_2(p)\,F(p)\} \quad (12)$$

From (12), the expression for $y(p)$ is found, namely:

Eqn (13) *

By substituting the expression for $y(p)$ from (13) in (9), one obtains, after some elementary calculations:

$$[1+kw_1(p)\,w_3(p)+kw_1(p)\,w_2(p)]\,x_{output}(p)$$
$$=kw_1(p)\,w_2(p)\,x_{input}(p)$$
$$-kw_1(p)\,w_2(p)\,w_{convert}(p)\,w_3(p)\,F(p)$$
$$+w_2(p)\,F(p)+kw_1(p)\,w_2(p)\,w_3(p)\,F(p)\ldots \quad (14)$$

Assume that the transfer function of the stabilizing device has been selected in such a way that the structure obtained assures stability where there exists an unlimited increase in the amplification factor, k. Dividing eqn (14) by k, and assuming that k tends to infinity, one obtains, after some simplifications:

$$[w_3(p)+w_2(p)]\,x_{output}(p)=w_2(p)\,x_{input}(p)$$
$$+[w_2(p)\,w_3(p)-w_2^2(p)\,w_{convert}(p)\,w_3(p)]\,F(p)\ldots \quad (15)$$

As is evident from (15), in order to eliminate the effect of interferences, the transfer function of the converting device should be selected from the condition:

$$w_2(p)\,w_3(p)-w_2^2(p)\,w_{convert}(p)\,w_3(p)=0 \quad (16)$$

or:

$$w_{convert}(p)=1/w_2(p)$$

The realization of a device with a transfer function (16) may be attained by methods of plotting structures that are stable in the face of an unlimited increase in the amplification factor (6), and it presents neither fundamental nor technical difficulties.

Generally speaking, the elimination of the influence of interferences, in the given case, could be accomplished by the method described by the author[7]. Naturally it is expedient to make use of the indicated method if there are no additional interferences at the system's input. If, at the system's input, there are interferences, in addition to the useful signal, then it is possible to show that the solution given here is more noise-proof. Let us convince ourselves of the accuracy of this affirmation.

It is assumed that, in place of the transfer function $kw_1(p)$, and in place of a stabilizing device with a transfer function

---

* *Eqn (13):*

$$Y(p)=\frac{kw_1(p)\,x_{input}(p)-kw_1(p)\,x_{output}(p)-kw_1(p)\,w_2(p)\,w_{convert}(p)\,w_3(p)\,F(p)}{1+kw_1(p)\,w_3(p)} \quad (13)$$

$w_3(p)$, which provides for stability in the face of an unlimited amplification factor k, a system having the form shown in *Figure 4* is realized.

The introduction of an amplifier, with a high amplification factor, which is encompassed by a stabilizing device with a transfer function w_3', depends on the necessity of providing stability to the entire system in the face of unlimited increase in k. If $w_2(p)$ has a power 'p' in the denominator that is greater than a 'fourth' one, then, as is shown in (6), it is possible to introduce several amplifiers with high amplification factors and realize a structure that would admit an unlimited increase in k without disturbing the stability.

As is clear from (7) in this case, if x_{input} contains no interferences, then an increase in the amplification factor k, up to rather high values, eliminates the effect of the F interferences.

Assume that the input signal contains an interference f_{input} to the extent that

$$x_{\text{input}} = x_{\text{input}\,u} + f_{\text{input}}$$

where $x_{\text{input}\,u}$ is the interference-free input signal.

A system of equations for the circuit in *Figure 4* is drawn up, for the case under examination. At the same time, instead of the part of the circuit surrounded by a dotted line in *Figure 4*, assuming that here k is a sufficiently large number, one should straightway insert $1/w_3'(p)$.

$$Y_1(p) = k[x_{\text{input}\,u}(p) + f_{\text{input}}(p) - x_{\text{output}}(p)] \ldots \quad (17)$$

$$Y_2(p) = Y_1(p)\frac{1}{w_3'(p)} = \frac{k}{w_3'(p)}[x_{\text{input}\,u}(p) + f_{\text{input}}(p) - x_{\text{output}}(p)] \quad (18)$$

$$x_{\text{output}}(p) = w_2(p)[Y_2(p) + F(p)] \quad (19)$$

Substituting the value $Y_2(p)$ from (18) in eqn (19), one obtains:

$$x_{\text{output}}(p) = \frac{w_2(p)\,x_{\text{input}\,u}(p)}{w_3'(p)} + w_2(p)\frac{kf_{\text{input}}(p)}{w_3'(p)} + w_2(p)\,F(p)\frac{kx_{\text{output}}(p)}{w_3'(p)}\,w_2(p)$$

or:

$$\left[1 + \frac{k}{w_3'(p)}\right]x_{\text{output}}(p) = \frac{kw_2(p)}{w_3'(p)}[x_{\text{input}\,u}(p) + f_{\text{input}}(p)] + w_2(p)\,F(p)$$

Where k tends to infinity, one obtains:

$$\frac{w_2(p)}{w_3'(p)}x_{\text{output}} = \frac{w_2(p)}{w_3'(p)}[x_{\text{input}\,u}(p) + f_{\text{input}}(p)]$$

or:

$$x_{\text{output}}(p) = x_{\text{input}\,u}(p) + f_{\text{input}}(p) \quad (20)$$

Consequently, one obtains at the output a magnitude that is equal to the ideal input plus the interference.

Consider, at this point, the size of the magnitude at the output, in the presence of interference at the system's input, and with the elimination of the effect of F interference by the above-mentioned method.

Keeping in mind that at the input of the system in *Figure 3*, along with the useful signal, there is an interference feed, one has the following system of equations (*Figure 3*)

Eqn (21) *

or, considering (11), one has:

Eqn (22) **

from which:

Eqn (23) †

Substituting the value of $y(p)$ from (23) in (9), and after some elementary calculations, one obtains:

Eqn (24) ††

On fulfilling the condition $w_2(p) = 1/w_3(p)$ and where k tends to infinity, one obtains:

$$x_{\text{output}}(p) = \frac{w_2(p)}{w_3(p) + w_2(p)}x_{\text{input}\,u}(p) + \frac{w_2(p)}{w_3(p) + w_2(p)}f_{\text{input}}(p) \quad (25)$$

By comparing the results expressed in eqn (20) and in eqn (25), one can draw the following conclusions. In the first case (eqn 20), the greater the amplification factor, the closer the output magnitude to the sum of the ideal input plus the full interference at the input. In the second case (eqn 25), the picture is different. Depending on the properties of the useful signal and of the interference, especially for those cases where the frequency properties of the interference, the parameters $w_3(p)$ may be selected in such a way as to reduce the interference, which enters at the input, together with the useful signal, to a minimum.

---

* *Eqn (21):* $$y(p) = kw_1(p)\{x_{\text{input}\,u}(p) + f_{\text{input}}(p) - x_{\text{output}}(p) - w_3(p)[y(p) + (x_{\text{output}}(p) - x'_{\text{output}}(p))\,w_{\text{convert}}(p)]\} \quad (21)$$

** *Eqn (22):* $$y(p) = kw_1(p)[x_{\text{input}\,u}(p) - f_{\text{input}}(p) - x_{\text{output}}(p) - w_3(p)\,y(p) - w_3(p)\,w_{\text{convert}}(p)\,w_2(p)\,F(p)] \quad (22)$$

† *Eqn (23):* $$y(p) = \frac{kw_1(p)[x_{\text{input}\,u}(p) + f_{\text{input}}(p) - x_{\text{output}}(p) - kw_1(p)\,w_2(p)\,w_3(p)\,w_{\text{convert}}(p)\,F]}{1 + kw_1(p)\,w_3(p)} \quad (23)$$

†† *Eqn (24):*

$$x_{\text{output}}(p) = \frac{kw_1(p)\,w_2(p)\,x_{\text{input}\,u}(p) + kw_1(p) + f_{\text{input}}(p) - kw_1(p)\,w_2^2(p)\,w_{\text{convert}}(p)\,w_3(p)\,F(p) + kw_1(p)\,w_2(p)\,w_3(p)\,F(p)}{1 + kw_1(p)\,w_3(p) + kw_1(p)\,w_2(p)} \quad (24)$$

A Change in the Object's Parameters Taking Place as a Result of a Change in the Operating Conditions or in the Internal Factors

This case pertains to plants, in which, in the course of operation, the parameters of the object itself may vary within a wide range. In such cases, the sensitivity factor, according to Bode[8], represents an essential qualitative index of the entire system. For the plants being considered here, the sensitivity factor may be expressed in the following manner.

Assume that the object's transfer function, as before, is designated by $w_2(p)$. The overall tranfer function of the entire system in relation to changes in the object indicated by $k(p)$, is expressed in the following way:

$$S^{k(p)}_{w_2(p)}=\frac{\dfrac{dk(p)}{k(p)}}{\dfrac{dw_2(p)}{w_2(p)}}=\frac{dk(p)}{dw_2(p)}\cdot\frac{w_2(p)}{k(p)} \tag{26}$$

In the general case, the smaller the magnitude of $S^{k}_{w2}(p)$, the less sensitive are the dynamic properties of the system, in its entirety, to changes in the plant's properties. For this case, the system is considered ideal or self-adjusting, if the magnitude $S^{k(p)}_{2(p)}$ does not depend on the characteristics of $w_2(p)$ or $S^{k(p)}_{w_2(p)}$ tends to 0.

The following proof is given. Structures that are stable in the face of an unlimited increase in the amplification factors, in which stability is achieved by the introduction of ideal derivatives and whose degree of ideality is determined by the magnitude of the amplification factor (7), belong to the category of self-adjusting systems in the sense indicated above. There is no question about that. In *Figure 5* one observes the structure of the regulating system of the type under consideration. The transfer function of the closed system will be written in the following form:

$$k(p)=\frac{\dfrac{k}{w_3(p)}w(p)}{1+\dfrac{k}{w_3(p)}w_2(p)} \tag{27}$$

Now the expression for the sensitivity is found. In conformity with (26):

$$S^{k(p)}_{w_2(p)}=\frac{\dfrac{k}{w_3(p)}\left[1+\dfrac{k}{w_3(p)}w_2(p)\right]-\left(\dfrac{k}{w_3(p)}\right)^2 w_2(p)}{\left[1+\dfrac{k}{w_3(p)}w_2(p)\right]^2}\cdot\frac{w_2(p)\left(1+\dfrac{kw_2(p)}{w_3(p)}\right)}{\dfrac{k}{w_3(p)}w_2(p)}$$

or, after simplification:

$$S^{k(p)}_{w_2(p)}=\frac{1}{1+\dfrac{k}{w_3(p)}w_2(p)} \tag{28}$$

where k tends to infinity, $S^{k(p)}_{w2(p)}$ tends to 0. In other words, in the sense indicated above, one obtains an ideal system.

Now consider the expression for sensitivity, if the structure belongs to the category of those that are stable in the face of an unlimited increase in the amplification factor, and where stability is achieved by the introduction of passive stabilizing devices.

As an example, one should consider the simplest type of such a system whose structure is shown in *Figure 6*.

The transfer function of the closed system in *Figure 6* is written as follows:

$$k(p)=\frac{kw_1(p)\,w_2(p)}{1+kw_1(p)\,w_3(p)+kw_1(p)\,w_2(p)} \tag{29}$$

The sensitivity, according to $w_2(p)$, is written:

Eqn (30) *

or, after simplification:

**

Where k tends to infinity, one has:

$$\lim_{k\to\infty} S^{k(p)}_{w_2(p)}=\frac{w_3(p)}{w_2(p)+w_3(p)} \tag{31}$$

Consequently, in the given case, even with sufficiently high amplification factors, a change in the parameters or characteristics of the plant exerts an influence on the dynamic properties of the system.

Consider some methods for perfecting the system's structure, with the object of reducing to the minimum the effect of the variation in the plant's characteristics on the system's dynamic properties, and in this manner, make the system self-adjusting in the above-determined sense.

Where external disturbances, which did not seem capable of measurement, acted on the object, in this case, too, it is expedient to introduce a plant model into the system, in order to obtain a self-adjusting system. A structural schematic diagram for the case under consideration is shown in *Figure 6*.

Keeping in mind the designations set forth in *Figure 6*, one writes:

Eqn (32) †

Here, $x'_{output}(p)$ is the representation for the output of the plant's model and x_{output} is the representation for the plant's output.

---

* *Eqn (30)*:

$$S^{k(p)}_{w_2(p)}=\frac{kw_1(p)\left[kw_1(p)\,w_3(p)+kw_1(p)\,w_2(p)+1\right]-kw_1(p)\,kw_1(p)\,w_2(p)}{\left[1+kw_1(p)\,w_3(p)+kw_1(p)\,w_2(p)\right]} \tag{30}$$

**

$$S^{k(p)}_{w_2(p)}=\frac{kw_1(p)\cdot kw_1(p)\,w_3(p)+kw_1(p)}{kw_1(p)\left[1+kw_1(p)\,w_3(p)+kw_1(p)\,w_2(p)\right]}\;\frac{kw_1(p)\,w_3(p)}{1+kw_1(p)\,w_3(p)+kw_1(p)\,w_2(p)}$$

† *Eqn (32)*:

$$Y(p)=kw_1(p)\left[x_{input}(p)-x_{output}-w_3(p)\,y-w_3(p)\left(x'_{output}(p)-x_{output}(p)\right)\right] \tag{32}$$

It is assumed that the model's characteristics remain invariable. Under those conditions, the difference $x'_{output}(p) - x_{output}(p)$ is equivalent to the disturbance that is conditional on the change in the plant's characteristics. Consequently

$$x'_{output}(p) - x_{output}(p) = CF(p) \tag{33}$$

C is a constant coefficient.
In this manner,

$$x_{output}(p) = w_2(p)\,y(p) = w'_2(p)\,y(p) + Cw'_2(p)\,F(p) \tag{34}$$

Substituting in eqn (32), instead of $x'_{output}(p) - x_{output}(p)$, the difference value from (33), one obtains:

Eqn (35) *

From the above, the expression for $y(p)$ is found:

$$y(p) = \frac{kw_1(p)\,x_{input}(p) - kw_1(p)\,k_{output}(p) - w_3(p)CF(p)}{1 + kw_1(p)\,w_3(p)} \tag{36}$$

Substituting the value for $y(p)$ from (36) in eqn (34), one obtains:

Eqn (37) **

or, determining $x_{output}(p)$ from (37), one obtains:

$$x_{output}(p) = \frac{kw_1(p)\,w'_2(p)\,x_{input} + w'_2(p)CF(p)}{1 + kw_1(p)\,w'_2(p) + kw_1(p)\,w_3(p)} \tag{38}$$

Where k tends to infinity:

$$x_{output}(p) = \frac{w_1(p)\,w'_2(p)\,x_{input}(p)}{w_1(p)\,w_3(p) + w_1(p)\,w'_2(p)} = \frac{w'_2(p)\,x_{input}(p)}{w_3(p) + w'_2(p)} \tag{39}$$

From (39) it is evident that the output magnitude does not depend on the change in parameters of the regulation plant.

Under the conditions where $w'(p)$ corresponds to the optimum operating circumstances, from the point of view of some qualitative criterion, the process in the system will be maintained automatically at these working conditions, independently of the plant's characteristic changes.

Thus, as a result of considering the three most interesting cases involving changes in the characteristics of the control plants–changes due to the effect if external disturbances, which could be measured, those due to external disturbances that did not appear to be capable of measurement, and those which resulted from plant characteristic changes in the course of operation that were independent of external disturbances–methods were suggested for designing structures that would provide for the independence of the plant's selected operating conditions from possible external and internal effects on it, and, consequently, the structures obtained proved to be self-adjusting system structures.

References

1. Fel'dbaum, A. A. On the use of computers in automatic systems. *Automat. Telemech., Moscow* No. 11 (1956)
2. Fel'dbaum, A. A. Automatic optimizer. *Automat. Telemech., Moscow* No. 8 (1958)
3. Doganovskii, S. A., and Fel'dbaum, A. A. Study of compensation for carrier thickness oscillations with the aid of an electron model. *Automat. Telemech., Moscow* No. 2 (1959)
4. Fel'dbaum, A. A. *Computers in Automatic Systems.* 1959. Moscow; Fizmatgiz
5. Lerner, A.Ya. Optimum high-speed control system by means of an automatic potentiometer. *Automat. Telemech., Moscow* No. 2 (1952)
6. Meerov, M. V. *Structure Synthesis of Highly-accurate Automatic Control Systems.* 1959. Moscow; Fizmatgiz
7. Meerov, M. V. On the structural noiseproof feature of one category of dynamic systems. *DAN, Proc. Acad. Sci.* No. 4 (1961)
8. *Chain Theory and Construction of Return-communication Feedback Amplifiers.* 1948. Moscow; Fizmatgiz

---

* *Eqn (35):*

$$Y(p) = kw_1(p)\,[x_{input}(p) - x_{output}(p)\,w_3(p)\,y(p) - w_3(p)\,CF(p)] \tag{35}$$

** *Eqn (37):*

$$x_{output} = \frac{w'_2(p)\,[kw_1(p)\,x_{input}(p) - kw_2(p)\,x_{output}(p) - w_3(p)\,kw_1(p)\,CF(p)}{1 + kw_1(p)\,w_3(p)} + w'_2(p)\,CF(p) \tag{37}$$

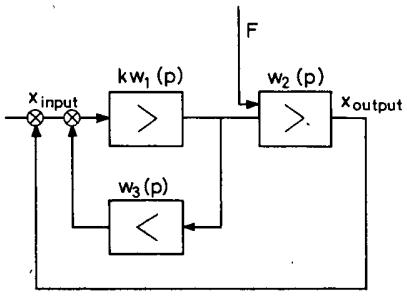


Figure 1.

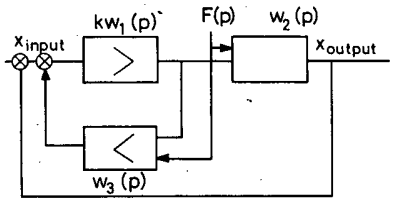


Figure 2.

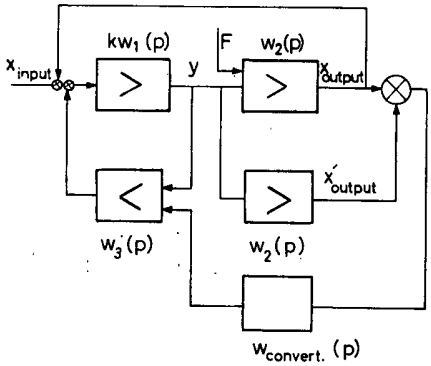


Figure 3.

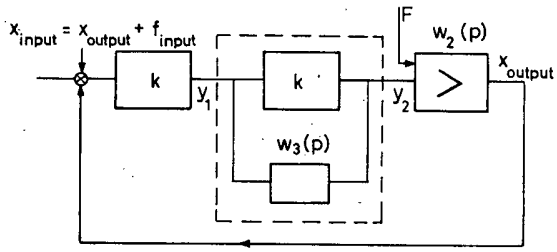


Figure 4.

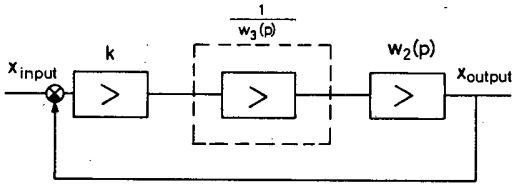


Figure 5.

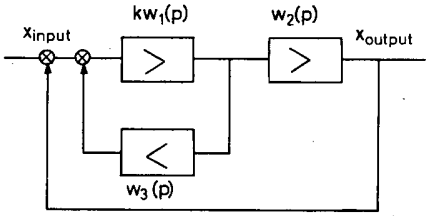


Figure 6.

A Method of Optimal Control Prediction

F. B. GULKO and B. YA. KOGAN

Introduction

In view of the increasing demands which are made on the quality of automatic control processes, more and more use is being made of optimal control systems, particularly of a wide class of time-optimal systems.

The development of time-optimal systems is at present badly hampered by the difficulty involved in designing the controlling part of the system, which, except for the simplest cases of second-order linear plants, involves the use of multivariable functional generators, or of complex boundary-problem computers[1].

Because of this, there has recently been a search for new approaches to the design of optimal control systems. In this connection mention should be made of the work of Coales and Noton[2], who proposed that search of the switching moment of the control action should be realized on the basis of high-speed examination of a family of phase trajectories (future behaviours of the plant), on the assumption that this switching will take place at some future moment. Chestnut, Sollecito and Troutman further, developed this principle[3], substituting for search of the switching moment a step-by step analysis of sections of the phase trajectory, also obtained at high speed, on the assumption that switching of the control action has taken place at the current moment of time; the actual switching is executed when the predicted trajectory passes through the origin of the coordinates. Characteristic features of the above works are: (*a*) Prediction by iterative computers of the set of future optimal behaviours (trajectories) of the plant, with verification of each trajectory to see whether it corresponds to the assigned boundary conditions; (*b*) the use in the control system of a logic for second-order plants, calculated for not more than one switching of the control action. The latter confines the possible applications of these methods to second-order plants, or to plants reducible to the second order, having no supplementary constraints on the coordinates. It is, however, possible to remove these constraints, at least for single-loop plants (or plants reducible to a single loop) consisting of first-order elements, linear or with monotonic non-linearities, by using the peculiarities of the structure of optimal processes in such systems. In this way it is possible to realize an optimal control system for a plant of the nth order, having an optimal controller for a plant of the (n-1)th order and a predictor. Applying the same principle in succession to plants of the (n-1)th, (n-2)th, ... orders, up to and including the second order, it is possible to construct an optimal control system for an nth order plant, the controlling part of which will consist of a set of predictors.

Construction of Optimal Control Systems by Successive Reduction of the Order and Prediction

Considered here are plants described by a system of differential equations of the form:

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, u) \\ \dot{x}_2 &= f_2(x_2, x_1) \\ &\cdots\cdots\cdots\cdots \\ \dot{x}_n &= f_n(x_n, x_{n-1}) \end{aligned} \qquad (1)$$

where $u = u(t)$ is the control action, while

$$|u(t)| \leq 1$$

All the functions f_i are assumed to be continuous and continuously differentiable with respect to x_i and x_{i-1}, while f_i is continuously differentiable with respect to u, and the partial derivatives $\partial f_i/\partial x_{i-1}$ and $\partial f_1/\partial u$ do not change sign throughout the domain of variation of the variables in question.

Moreover, on some x_k, there can be imposed constraints of the form:

$$|x_k| \leq \overline{x_k}$$

specifying the permissible domain of states of the system in the phase space. The problem is to synthetize a control system which effects the time-optimal shift of plant (1) from any initial state to any assigned equilibrium state.

To solve this problem, use will be made of a property of the structure of optimal processes in plants of type (1), namely that the trajectory of the optimal process consists of successive sections, on each of which the control corresponding to it coincides with the optimal control for a type (1) system having an order lower by a unity than the initial one. For example, if by some means it is possible to ensure, for a (n-1)th order plant [without the last element in (1)], a control action which in minimal time imparts to the coordinate x_{n-1}, an extremal value (taking into account the imposed constraints), and which then, in minimal time, transfers the plant to the assigned state (with respect to the $n-1$ coordinate), and if, moreover, the coordinate x_n reaches the assigned value at the final moment, then the control action and corresponding trajectory of the whole system (1) are time-optimal. The proof of this is given by Gulko[4] (for the case when constraints of the $\overline{x_k}$ type are lacking), where it is shown that such a control action satisfies Pontryagin's Maximum Principle[5].

Figure 1 is the block diagram of an optimal control system based on these principles. The scheme consists of three main parts: the plant itself with an optimal controller [for the (n-1)th order], the predictor (P), and the logical gate (L).

The optimal controller in the plant assures optimal motion of the (n-1)th order plant towards the value of the coordinate x_{n-1} assigned by the logical gate L. The predictor is a model of the plant together with controller optimal for the $n-1$ coordinate, the setting of which agrees with the assigned value of the coordinate x_{n-1}, operating iteratively at high speed. Obtaining, at the beginning of every cycle of the solution, data on the state

of the plant (the values of its current coordinates), the predictor computes the value which will be reached by the coordinate x_n if, starting from a given moment, the truncated (n-1) order system is brought to assigned equilibrium state in minimal time.

The output signal of the logical gate u_L is determined by the mismatch between the assigned value of the nth coordinate $x_{n\,\mathrm{spec}}$ and its predicted value $x_{n\,\mathrm{pred}}$, from the equation:

$$u_L \begin{cases} \bar{x}^*_{n-1} \operatorname{sign}(x_{n\,\mathrm{pred}} - x_{n\,\mathrm{spec}}) \cdot \operatorname{sign} \dfrac{\partial f_n}{\partial x_{n-1}} \\ \qquad \text{when } x_{n\,\mathrm{pred}} \neq x_{n\,\mathrm{spec}} \\ x_{n-1\,\mathrm{spec}} \qquad \text{when } x_{n\,\mathrm{pred}} = x_{n\,\mathrm{spec}} \end{cases} \tag{2}$$

With input of the specification $x_{n\,\mathrm{spec}}$ a mismatch arises between $x_{n\,\mathrm{spec}}$ and $x_{n\,\mathrm{pred}}$, as a result of which the logical gate L, in accordance with (2), sends to the optimal controller a specification for the variation of the coordinate x_{n-1}, (taking into account the sign of the mismatch). Moreover, the predictor continuously calculates the value which will be taken by the coordinate x_n if, at the given instant, the setting of the optimal controller is switched from $\bar{x}_{n-1}$ to $x_{n-1\,\mathrm{spec}}$. As soon as the value of $x_{n\,\mathrm{pred}}$ reaches $x_{n\,\mathrm{spec}}$, the logical gate will bring about an actual change of the setting of the controller, after which the system will adopt the specified position, under the influence of the optimal controller. The predicted value $x_{n\,\mathrm{pred}}$ remains unaltered over this interval of time.

The chain of reasoning employed for the synthesis of an optimal system of the nth order can be used to synthetize an (n-1)th order optimal controller for plant and predictor; that is, recourse is made to a system with an (n-2)th order optimal regulator and two predictors for the coordinates x_n and x_{n-1}. Naturally, in this case, the predictor which works out the future value of the coordinate x_{n-1}, and which itself forms part of the predictor for the coordinate x_n, must operate at a higher speed than the latter. Applying this method successively a further $n-3$ times, one arrives at an optimal control system containing $n-1$ predictors ($P_1, P_2, \ldots, P_{n-1}$) with their corresponding logical gates ($L_1, L_2, \ldots, L_{n-1}$), but containing no other optimal controllers (*Figure 2*). It is a characteristic feature of this system that the optimal nature of the calculated trajectories in any of the predictors is ensured by the presence in the make-up of any of them of other predictors which calculate the motion of a successively abbreviated number of elements at ever-increasing speed. *Figure 3* is a block diagram illustrating the method of synthesis of the predictors.

To solve a tracking problem by the method described, recourse must be made to error equations, as was done by Coales and Noton[2], and by Chestnut *et al.*[3].

Optimal Control of a Fourth-order Plant

To illustrate the method, *Figure 4* shows the example of a time-optimal control system for a fourth-order plant consisting of four integrating elements. The system contains three predictors: P_1, P_2 and P_3. Let there be supplied to the system at some moment a 'specification' concerning the coordinate x_4. If at this moment the state of the system is such that, with optimal alteration of the coordinate x_3 to the specified value corresponding to an equilibrium state (i.e., to zero), the coordinate x_4 does not reach the specified magnitude, then as a result of the mismatch between $x_{4\,\mathrm{spec}}$ and $x_{4\,\mathrm{pred}}$ the logical gate L_3 gives a signal u_{L3}, corresponding to the limit permissible value of the coordinate x_3, with the appropriate sign determined by the direction of the 'acceleration' of the coordinate x_4. If there is a mismatch between u_{L3} and $x_{3\,\mathrm{pred}}$, logical gate L_2 gives a 'specification' U_{L2} for the variation of the coordinate x_2, and logical gate L_1, under the same conditions, gives a specification u_{L1} for a variation of the coordinate x_1 which switches the control relay. After this the system begins to 'accelerate' with maximum rapidity in the required direction, and at some moment t_1 the value of $x_{4\,\mathrm{pred}}$ becomes equal to the given value $x_{4\,\mathrm{spec}}$. Starting from this moment, the coordinate x_3 must be zeroed with optimal rapidity, so logical gate L_3 changes the command signal u_{L3} to zero. In the same way, logical gates L_2 and L_1 change the signs of u_{L2} and u_{L1}, and the first switching of the command relay takes place, after which the system starts to 'brake' with respect to the coordinate x_4. During the rest of the process, $x_{4\,\mathrm{pred}}$ will equal $x_{4\,\mathrm{spec}}$. When the value of $x_{3\,\mathrm{pred}}$ reaches zero, gate L_2 issues a command for optimal change of coordinate x_2 to zero. This command enters the relay via gate L_1, and effects the second switching. After this, $x_{3\,\mathrm{pred}}$ remains equal to zero. The third switching occurs in the same way, when $x_{2\,\mathrm{pred}} = 0$. This last interval ends when x_1 reaches equilibrium value (zero). Thus by the end of the process $x_3 = x_2 = x_1 = 0$ and $x_4 = x_{4\,\mathrm{spec}}$. If, in the course of the process, the predicted value of some one of the coordinates, for example x_3, reaches a magnitude equal to the limit value prescribed for it by logical gate L_3, the specification for x_2 will be switched to zero by the gate L_2, and the system will start 'braking' with respect to the coordinate x_3. At the end of this 'braking' process, x_3 reaches its permissible value, and maintains it until the predicted value of x_4 reaches the specified value. In this case the process will consist of a greater number of switchings, as follows from the theory of optimal control[6]. The results of simulating control processes by the proposed method are shown in the oscillograms of *Figure 5* (*a*), (*b*), and (*c*), which show processes for the cases of no constraints on the phase coordinates, and the application of constraints on the third coordinate and the second and third coordinates together. Oscillograms of the outputs of the predictors P_3 ($x_{4\,\mathrm{pred}}$) and P_2 ($x_{3\,\mathrm{pred}}$) are also given.

If it is possible to realize optimal control of the first k elements of the plant ($k = 2, 3, \ldots$) in some other way, the number of predictors can be reduced by $k-1$. The first of the remaining predictors must include a model of the corresponding optimal controller, the second a model of the first predictor, and so on, as has been described for the general case.

Some Features of the Design of Iterative Analogue Predictors

A predictor is a high-speed iterative computer, operating with acceleration of the processes (with respect to the plant). Because of the need for a high iteration rate, while the requirements for accuracy are relatively low, the use of analogue principles in the design of the predictors is most expedient.

The iteration rate (periodicity) is selected from considerations of the increment of the predicted quantity permissible, for reasons of accuracy, in a cycle of the solution. The time scale is chosen with references to the iteration rate and the duration of the processes in the plant to be predicted by the device.

* If no constraint x_{n-1} is given, then the greatest value of the coordinate x_{n-1} that can be physically represented is introduced into the logic block in its place.

A predictor usually consists of the following main units: an analogue of the relay device supplying the control action; iterative analogue computing elements (linear or non-linear), with a wide pass band; a memory element, which stores information between cycles of the solution, and, finally, a control system, which provides the necessary sequence of switching operations.

The analogue relay element is usually a flip-flop or a computing amplifier with a restrictor in the feedback circuit or on the output.

The amplifiers for the computing elements must ensure the desired accuracy of operation, and must have low drift. For this purpose, according to Polonnikov[7], the most suitable is a direct current amplifier with a zero drift compensation network based on the ideas of Prinz. *Figure 6* gives the structural scheme of a scale computing element and integrator. Here in the RESET cycle all the switches are closed, and because of the full negative feedback the capacitor C_k is charged up to the drift voltage at the amplifier output. In the 'solution' cycle the switches are opened, and the compensating voltage across the capacitor C_k is connected in series with the voltage at the summing point. Ordinary switches of the bridge type are used.

The memory element stores the solution at the end of an operational cycle for the duration of the RESET period of the next solution cycle. It can be constructed from a combination of a computer amplifier, memory capacitors and switches, for example.

The control system produces cycle pulses which control the mode of operation of the switch, and can be constructed using conventional multivibrators.

Conclusions

The theoretical possibility of constructing time-optimal control systems for nth order plants has been demonstrated, using a set of predictors as the optimal controllers.

Such optimal control systems are synthetized according to a definite pattern, on the basis of a mathematical analogue of the plant. Elaborate calculations are not required, and the setting of the control system is simplified.

The predictors used in the control system can also function as 'advisors' to the operators in the case of manual control. For third-order and fourth-order plants, the method described can be realized with the aid of a comparatively simple apparatus, which can be built with the technical means now available.

The extension of the method to other, more complex, optimal control problems will require further investigation of the structural features of optimal processes, and the development of very high-speed and reliable means of mathematical simulation.

References

1 Feldbaum, A. A. *Computers in Automatic Systems.* Ch. 13. 1959. Moscow; Fizmatgiz

2 Coales, J. F., and Noton, A. R. M. An on-off servomechanism with predicted changeover. *Proc. Instn elect. Engrs* Pt B, 103, No. 10 (July 1956), 449–462

3 Chestnut, H., Sollecito, W. E., and Troutman, P. H. Predictive control system application. *Appl. and Ind.* 55 (July 1961), 128–134

4 Gulko, F. B. A special feature of the structure of optimal processes. *Automat. Telemech.* (in press)

5 Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., and Mishchenko, E. F. *A Mathematical Theory of Optimal Processes.* 1961. Moscow; Fizmatgiz

6 Lerner, A. Ya. Maximal quick response of automatic control systems. *Automat. Telemech.* 15, No. 6 (1954), 461–477

7 Polonnikov, D. F. *Computer Amplifiers for Iterative Simulators.* 1961. Moscow; 2nd All-Union Seminar-Conference on the Theory and Methods of Mathematical Simulation Moscow 1961

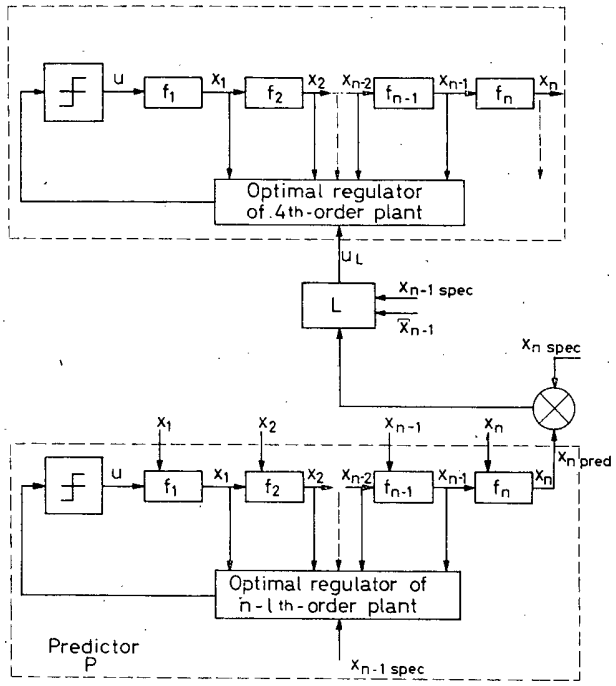


Figure 1. Block diagram of an automatic control system with prediction for one coordinate

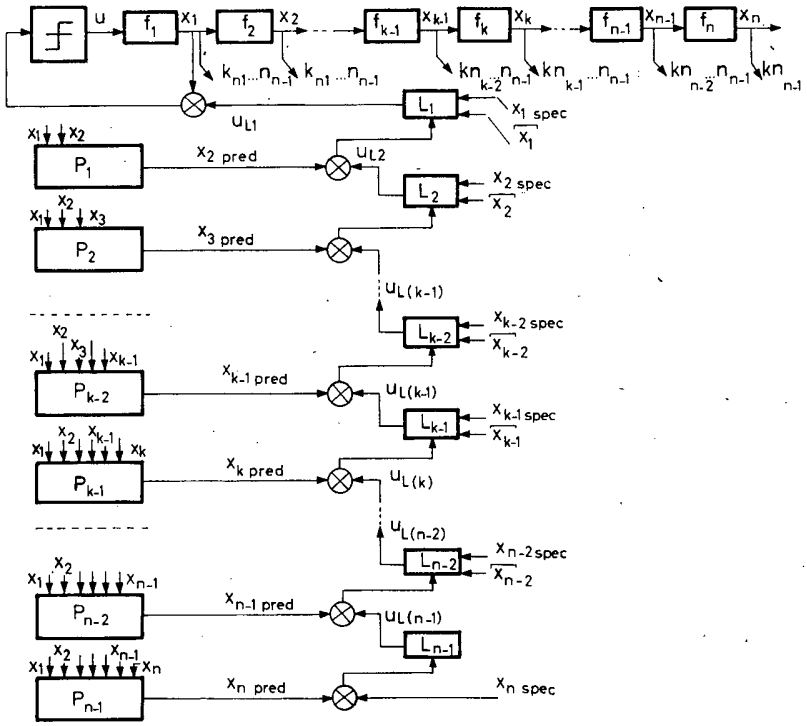


Figure 2. Block diagram of optimal control with prediction for $n-1$ coordinate (the general case)

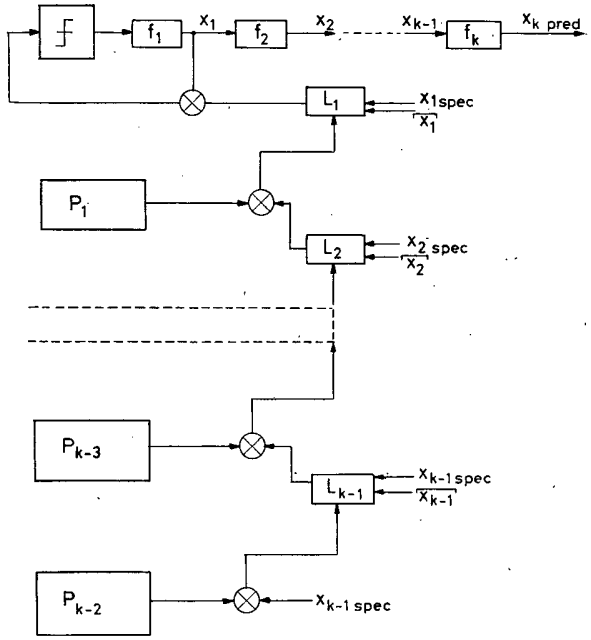


Figure 3. Block diagram of the $(k-1)$th predictor

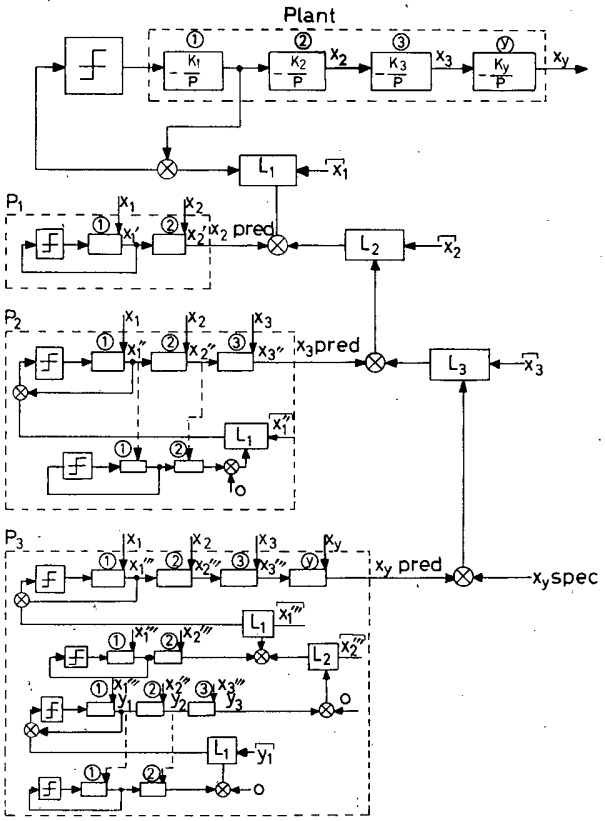


Figure 4. Automatic control system for fourth-degree plant

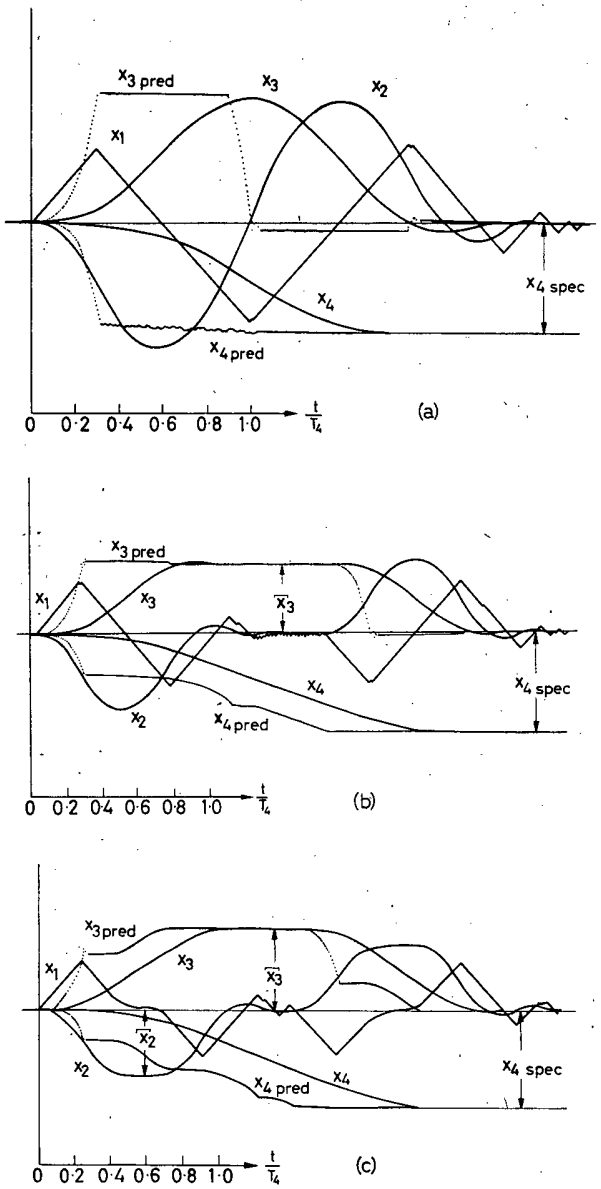


Figure 5. Oscillograms of transient processes in the optimal control system of a fourth-degree plant:

(a) with no constraint on the phase coordinates; (b) with a constraint on the coordinate x_3; and (c) with a constraint on the coordinates x_2 and x_3

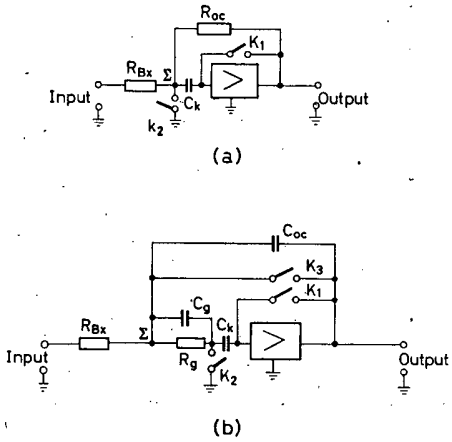


Figure 6. Zero drift compensation circuit of computing elements. (a) Scale computing element, and (b) integrator

On the Theory of Self-tuning Systems with a Search of Gradient by the Method of Auxiliary Operator

I. E. KAZAKOV and L. G. EVLANOV

Structure and Equations of a Self-tuning System

In many cases important in practice, automatic control systems may be represented in the form of a generalized system illustrated in *Figure 1*. The object of control is charaterized by an operator of a given structure $A(\eta)$, where η is a group of parameters for which *a priori* information is lacking. The system of control is described by an operator $B(\xi)$ which depends on the group of parameters ξ_i $(i = 1, 2, \ldots, n)$ which may be tuned. In actual systems, the aggregate of values of each parameter ξ_i forms a finite multitude Ξ_i. The input signals of the system are $X(t)$, the useful random signal, and $Z(t)$, $U(t)$, random-disturbances.

The equations of the automatic control system are as follows:

$$\begin{aligned} Y &= A(\eta)[V+U] \\ V &= B(\ )\varepsilon \\ \varepsilon &= X+Z-Y \end{aligned} \tag{1}$$

In order to assure high quality functioning of the automatic control system it is necessary to achieve tuning of parameters of the operator $B(\xi)$ in the presence of variation of the characteristics of the input useful signal $X(t)$, of the characteristics of disturbances $Z(t)$, $U(t)$, and also in the presence of variation of parameters η of the operator of the object of control.

In order to construct a circuit for self-tuning, an index of quality I of the automatic control system is introduced. The index of quality I is a function, or in the general case it is a functional of tuned parameters. Ordinarily the index of quality I is computed on the basis of error ε of the system:

$$I = Nf(\varepsilon, \xi) \tag{2}$$

where N is an operator or a functional, $f(\varepsilon,\xi)$ is a function of the error of the system depending upon the error ε and the tuned parameters ξ.

In order to tune the parameters of the system use is made of the broad possibilities offered by the method of steepest descending slope or gradient, a discussion of which is considered by Feldbaum[1]. Applying this method for tuning parameters ξ one has:

$$\dot{\bar{\xi}} = \lambda \operatorname{grad} I \tag{3}$$

where λ is the scalar multiplier, and $\dot{\bar{\xi}}$ is a vector function of the velocities of tuned parameters. In accordance with the gradient method the self-tuning system assures the tuning of parameters ξ for the optimal value of index or quality I_0. In the general case

$$I_0 = \inf_{\xi_i \in \Delta_i} I(\xi) \quad \text{or} \quad I_0 = \sup_{\xi_i \in \Delta_i} I(\xi) \tag{4}$$

In the particular case when the lower (upper) boundary of the multitude ξ_i is attained within Ξ_i,

$$I_0 = \operatorname{extremum} I(\xi) \tag{5}$$

For a complete description of the circuit for self-tuning it is necessary to determine the method of computation of the components of the gradient from the quality index for the tuned parameters. In the given investigation a method is applied which, in the following is termed the method of an auxiliary operator. Its essence consists of the following.

If the information on operators $B(\xi)$ and $A(\eta)$ is known *a priori*, it is possible to construct a certain auxiliary operator $C(\xi,\eta)$ whose application to the error of the tracking system makes it possible to compute the components of the gradient vector.

The derivative $\partial I/\partial \xi_i$ is computed by the direct differentiation of the expression (2) assuming that the operators N and differentiations with respect to ξ_i are commutative.

$$\frac{\partial I}{\partial \xi_i} = N\frac{\partial f(\varepsilon,\xi)}{\partial \varepsilon}\cdot\frac{\partial \varepsilon}{\partial \xi_i} + N\frac{\partial f_i(\varepsilon,\xi)}{\partial \xi_i} \tag{6}$$

The derivative $\partial \varepsilon/\partial \xi_i$ will be calculated by differentiating the system of eqns (1). The derivative of the error ε with respect to ξ_i is equal to

$$\frac{\partial \varepsilon}{\partial \xi_i} = -\frac{\partial Y}{\partial \xi_i} \tag{7}$$

since the input signals $X(t)$, $Z(t)$ do not depend upon ξ_i. The derivatives of the output signal are computed:

$$\frac{\partial Y}{\partial \xi_i} = A(\eta)\frac{\partial B(\xi)}{\partial \xi_i}\varepsilon + A(\eta)B(\xi)\frac{\partial \varepsilon}{\partial \xi_i} \tag{8}$$

Excluding from (7) and (8) $\partial Y/\partial \xi_i$ and transforming, one obtains:

$$\frac{\partial \varepsilon}{\partial \xi_i} = -[1+A(\eta)B(\xi)]^{-1}A(\eta)\frac{\partial B(\xi)}{\partial \xi_i} \tag{9}$$

Introducing the designation

$$C_i = [1+A(\eta)B(\xi)]^{-1}A(\eta)\frac{\partial B(\xi)}{\partial \xi_i} \tag{10}$$

one writes:

$$\frac{\partial \varepsilon}{\partial \xi_i} = -C_i(\eta,\xi)\varepsilon \tag{11}$$

or

$$\operatorname{grad}\varepsilon = -\bar{C}(\eta,\xi)\varepsilon \tag{12}$$

where $\overline{C}(\eta,\xi)$ is an auxiliary operator-vector which is completely determined by the operators $A(\eta)$, $B(\xi)$. Thus, the gradient of the quality index for tuned parameters is determined by eqns(6) and (11).

The method of auxiliary operator requires an *a priori* knowledge of information on the system, and this somewhat restricts its generality. However, there exists in technology an area of applicability of the method inasmuch as the predominant majority of created automatic control systems can be described mathematically.

The advantages of the method are the absence of trial load changes and the possibility of accelerating and simplifying the process of computation of the gradient components. In self-tuning systems with a search of gradient by the method of trial load changes, *a priori* information on the object, other than the knowledge of the band pass of the system, is not required. This permits a correct selection of the frequency of the trial load changes and constitutes the advantage of this method. However, its basic shortcoming is the limited quick response imposed by the finite band pass width of the system. In the considered method the band pass of the mathematical model of the system (operator C) may be artificially broadened by changing the time scale of the solution. The possibility of simplifying the process of computation is based on the substitution for a complex operator C of an approximate and simpler expression.

The auxiliary operator $\overline{C}(\eta,\xi)$ depends upon the parameters of the object and the system of control. A typical case is one of absence of *a priori* information on parameters η. Information on parameters of the object may be obtained on the basis of application of a tracking system, certain aspects of whose application were considered by Margolis and Leondes[2,3].

The structure of the operator of model $A(\zeta)$ is based on the utilization of *a priori* information on the object. The aggregate of parameters ζ of the operator of the model is tuned for the value η. The circuit of the tracking model is constructed quite analogously to the circuit for tuning. Introducing an index of approximation J of parameters ζ into parameters η,

$$J = L\phi(\varepsilon_1) \tag{13}$$

where L is an operator for computing the index J, and $\phi(\varepsilon_i)$ is a function of the error. The error is determined by the relationship

$$\varepsilon_1 = Y_M(t) - Y(t) \tag{14}$$

Here $Y_M(t)$ is an output signal of the model determined by the expression

$$Y_M(t) = A(\zeta)V \tag{15}$$

The change of the parameters of the model is carried out by the method of steepest descending slope:

$$\dot{\overline{\zeta}} = \lambda_1 \operatorname{grad} J \tag{16}$$

where λ_1 is a scalar multiplier, and $\dot{\overline{\zeta}}$ is a vector function of the velocities of the tuned parameters of the model.

In order to determine the components of the gradient one applies the method of auxiliary operator:

$$\frac{\partial J}{\partial \zeta_i} = L\frac{\partial \phi(\varepsilon_1)}{\partial \varepsilon_1}\cdot\frac{\partial \varepsilon_1}{\partial \zeta_i} \tag{17}$$

Differentiating the relationship (14) with respect to ζ_i, one has:

$$\frac{\partial \varepsilon_1}{\partial \zeta_i} = \frac{\partial Y_M}{\partial \zeta_i} = \frac{\partial}{\partial \zeta_i}A(\zeta)V = \frac{\partial A(\zeta)}{\partial \zeta_i}V \tag{18}$$

hence it follows that the auxiliary operator in a given case is an operator-vector $\overline{G}(\zeta)$ with components

$$G_i(\zeta) = \frac{\partial A(\zeta)}{\partial \zeta_i} \tag{19}$$

Thus

$$\operatorname{grad} J = L\left\{\frac{\partial \phi(\varepsilon_i)}{\partial \varepsilon_i}\overline{G}(\zeta)V\right\} \tag{20}$$

Equations (13), (14), (15), (16) and (20) describe the operation of the tracking model. A useful output of the circuit of the model is the aggregate of parameters of model ζ. For ideal operation of the model $\zeta \equiv \eta$. An actual model assures the attainment of parameters ζ close to values η, and therefore, strictly speaking, in the operator $\overline{C}$ it is necessary to replace parameters η by ζ.

The complete structural diagram of the self-tuning system in accordance with eqns (1), (3), (6), (11), (14), (15), (16) and (20) is presented in *Figure 2*. The schematic diagram was proposed by Evlanov.

The structure of the self-tuning system contains three circuits: the basic circuit of the system, the circuit of the tracking model, and the circuit of tuning of parameters. The circuit of the tracking model assures the reception of information on the parameters of the operator of the object. In the following the operation of the circuit of the tracking model is assumed to be ideal, that is, $\zeta \equiv \eta$. The circuit for tuning the parameters assures the tuning of parameters of the control system in accordance with the given optimal value of the quality index of the system.

Investigation of a Self-tuning System a Quasi-stationary Regime

A typical regime of operation of a self-tuning system is the case of a change of parameters η of the operator $A(\eta)$ of the object and the characteristics of external random disturbances X, Z, U which are slow compared with the duration of transitional processes in the basic circuit of the system. In this case it is permissible to consider the circuits of tuning parameters and the tracking model on one hand, and the basic circuit on the other hand, as being autonomous, since the tuned parameters ξ and parameters η may be considered as constant during the time of process control in the basic circuit. It is also assumed that the tracking model carries out its functions in an ideal manner. Under these conditions the process of self-tuning of parameters ξ of operator $B(\xi)$ is investigated in the vicinity of extremum of the quality index I.

The presence of extremum in the quality index I of the system with respect to all or several of the tuned parameters is an important property of the self-tuning systems which permits them to be tuned for an optimal regime. If the error of the system ε or another characteristics does not possess extremal properties, then it is possible to construct an extremal quality index by artificial means depending upon the direction of aiming of the automat. This will be shown below by an example of a typical tracking system. For the time being, however, it is assumed that the quality index I possesses extremal properties.

The random error ε of the basic circuit can be expressed in the form

$$\varepsilon = m_\varepsilon + \varepsilon^0 \tag{21}$$

where m_ε is the mathematical expectation, and ε^0 is the centring component of magnitude ε. In the function of the error $f(\varepsilon,\xi)$ we shall also factor out the mathematical expectation

$$f(\varepsilon,\xi) = Mf(\varepsilon,\xi) + f^0(\varepsilon,\xi) \tag{22}$$

where M is the operation of mathematical expectation, $f^0(\varepsilon,\xi)$ is the random centred component.

The quality index of control I introduced previously may now be presented as:

$$I^* = NI^* + Nf^0(\varepsilon\xi) \tag{23}$$

where the designation I^* is introduced for the statistical quality index of control

$$I^* = Mf(\varepsilon,\xi) \tag{24}$$

Computing the components of the gradient of the quality index of control by parameters ξ_i, one obtains:

$$\frac{\partial I}{\partial \xi_i} = N\frac{\partial I^*}{\partial \xi_i} + N\frac{\partial f^0}{\partial m_\varepsilon}\cdot\frac{\partial m_\varepsilon}{\partial \xi_i} + N\frac{\partial f^0}{\partial \varepsilon^0}\cdot\frac{\partial \varepsilon^0}{\partial \xi_i} + N\frac{\partial f^0}{\partial \xi_i} \tag{25}$$

Representing the statistical quality index I^* of control in the vicinity of the investigated extremum by a quadratic form in terms of deviations $u_i = \xi_i - \xi_{i0}$ of parameters ξ_i from the optimal values ξ_0, and considering that

$$\left[\frac{\partial I^*}{\partial \xi_i}\right]_{\xi_i=\xi_{i0}} = 0$$

at the point of extremum, we shall obtain for the current values of $\partial I^*/\partial \xi_i$ the expressions:

$$\frac{\partial I^*}{\partial \xi_i} = \sum_{j=1}^{n}\frac{1}{2}\left[\frac{\partial^2 I^*}{\partial \xi_i \partial \xi_j}\right]_0 u_j \tag{26}$$

Differentiating expressions (24) twice with respect to parameters ξ_i, ξ_j and utilizing a system of equations of the basic circuit of control for optimal parameters ξ_{i0} of operator $B(\xi)$, one computes the coefficients

$$\left[\frac{\partial^2 I^*}{\partial \xi_i \partial \xi_j}\right]_0$$

in the form:

$$\begin{aligned}\left[\frac{\partial I^*}{\partial \xi_i \partial \xi_j}\right] = M\Bigg\{&\frac{\partial^2 f(\varepsilon_0,\xi_0)}{\partial \varepsilon_0^2}(C_{j0}\varepsilon_0)(C_{i0}\varepsilon_0) \\ &+\frac{\partial f(\varepsilon_0,\xi_0)}{\partial \varepsilon_0}(C_{j0}C_{i0}\varepsilon_0) + \frac{\partial^2 f(\varepsilon_0,\xi_0)}{\partial \varepsilon\,\partial \xi_j}(C_{i0}\varepsilon_0) \\ &+\frac{\partial^2 f(\varepsilon_0,\xi_0)}{\partial \xi\,\partial \xi_i}(C_{j0}\varepsilon_0) + \frac{\partial^2 f(\varepsilon_0,\ 0)}{\partial \xi_i \partial \xi_j}\Bigg\}\end{aligned} \tag{27}$$

where $C_{i0}(\xi_0,\eta)$ are the auxiliary operators (10) for optimal values of parameters ξ_{i0}.

Introduce the designations:

$$\frac{1}{2}\left[\frac{\partial^2 I^*}{\partial \xi_i \partial \xi_j}\right]_0 = a_{ij} \tag{28}$$

Taking into account also that

$$\frac{\partial m_\varepsilon}{\partial \xi_i} = -C_1 m_\varepsilon,\quad \frac{\partial \varepsilon^0}{\partial \xi_i} = -C_i\varepsilon^0 \tag{29}$$

the formula (25) is written for the components of the gradient of the magnitude I in the form:

$$\frac{\partial I}{\partial \xi_i} = N\sum_{j=1}^{n} a_{ij}u_j - N\frac{\partial f^0}{\partial m_\varepsilon}C_i m_\varepsilon - N\frac{\partial f^0}{\partial \varepsilon^0}C_i\varepsilon^0 + N\frac{\partial f^0}{\partial \xi_i} \tag{30}$$

Substituting the expression (30) into formula (3), one obtains a system of equations of the circuits of tuning of parameters ξ_i in a scalar form:

$$\dot{\xi}_i = \lambda N\sum_{j=1}^{n} a_{ij}u_j - \lambda N\frac{\partial f^0}{\partial m_\varepsilon}C_i m_\varepsilon - \lambda N\frac{\partial f^0}{\partial \varepsilon^0}C_i\varepsilon^0 + \lambda N\frac{\partial f^0}{\partial \xi_i} \tag{31}$$

From this one obtains a system of linear equations for the determination of mathematical expectations of deviations m_{u_i} of tuned parameters from the optimal values:

$$\dot{m}_{u_i} - \lambda N\sum_{j=1}^{n} a_{ij}u_j = -\dot{\xi}_{i0} \tag{32}$$

In order to determine random components of deviations of tuned parameters u_i^0 one obtains the following system of linear equations:

$$\begin{aligned}\dot{u}_i^0 - \lambda N\sum_{j=1}^{n} u_j^0 a_{ij} = &-\lambda N\left[\frac{\partial f^0}{\partial m_\varepsilon}\right]C_{i_0}m_{\varepsilon 0} \\ &-\lambda N\left[\frac{\partial f^0}{\partial \varepsilon^0}\right]_0 C_{i_0}\varepsilon_0^0 + \lambda N\frac{\partial f^0}{\partial \xi_i}\end{aligned} \tag{33}$$

An analysis of approximate linear equations (32) makes it possible to evaluate the stability of the process and to determine the systematic errors of self-tuning of parameters ξ_i. In particular, if the basic circuit of control is stationary and possesses astatism of the kth order, then for stationary random disturbances Z and U, and for an additive component of the useful signal X in the form of a polynomial of the kth order, the left-hand parts of eqns (32) are stationary. In this widely encountered case the stability of self-tuning of the parameters is characterized by properties of characteristic equation. In this case the investigation of stability is carried out by ordinary means. In the general case the systematic components of the errors of parameters are computed by equations:

$$m_{u_i}(t) = -\sum_{j=1}^{n}\int_0^t g_{ij}(t,\tau)\,\dot{\xi}_{j0}(\tau)\,d\tau \tag{34}$$

where $g_{ij}(t,\tau)$ are the weight functions of the system of eqns (32). If ξ_{i0} = const., then the systematic values of errors of tuning of parameters $m_{u_i} = 0$. Dispersions of the errors of parameters are determined on the basis of the system of eqns (33) by applying the theory of transformation of random functions[4].

From the analysis of stability, duration of transitional processes of tuning, and evaluation of the precision, one chooses the coefficient λ and also other characteristics of the circuits of tuning.

The final evaluation of mathematical expectation of the

error in the basic circuit of the system under the action of circuits of self-tuning is obtained by the formula:

$$m_\varepsilon = m_{\varepsilon_0} + \sum_{i=1}^{n} m_{\varepsilon_i} \tag{35}$$

where m_{ε_0} is the mathematical expectation of the error of control ε for an optimal value of parameters.

The magnitudes m_{ε_i} are determined by the expressions:

$$m_{\varepsilon_i} = C_{i_0}(\xi_0)[c_i m_{u_i}]$$

where $C_{i_0}(\xi_0)$ are the auxiliary operators for optimal values of parameters ξ_{i_0} and the magnitudes b_i are equal to

$$b_i = \left[\frac{\partial B(\xi)}{\partial \xi}\right]_0 m_{\varepsilon_0} \tag{36}$$

The evaluation of dispersion of the error in the basic circuit is computed by the formula:

$$D_\varepsilon = D_{\varepsilon_0} + 2\sum_{i=1}^{n} k_{\varepsilon_0 \varepsilon_i} + \sum_{i,j=1}^{n} k_{\varepsilon_i \varepsilon_j} \tag{37}$$

where D_{ε_0} is the dispersion for optimal values of parameters ξ_{i_0}, $K_{\varepsilon_0 \varepsilon_i}$, $K_{\varepsilon_i \varepsilon_j}$ are the coefficients of correlation of random components of the error of control ε_i^0, and the magnitudes ε_i^0 are equal to

$$\varepsilon_i^0 = -u_i^0 [C_{i_0}(\xi_0) m_{\varepsilon_0}] \tag{38}$$

Linear Tracking System with One Tuned Parameter

The application of the method to a linear tracking system, with one tuned parameter, is now described. In tracking systems, as a rule, the index of quality of control is assumed to be the second initial moment of error ε. This magnitude does not possess extremal properties with respect to parameters ξ corresponding to the change of input random actions X, Z, U.

Now consider an example of a tracking system having the following characteristics: $A(\eta) = \frac{\eta_1}{D}$, $B(\xi) = \xi_1$, $X = at$, $U = 0$,

$$m_z = 0, s_z = \frac{D_z}{\pi} \frac{\beta}{(\omega^2 + \beta^2)}$$

and values of parameters given by $\eta_1 = 10$, $a = 0, 1$, $D_z = 10^{-4}$, $\beta = 100$. The second initial moment of error ε in a stabilized regime is equal to:

$$\alpha_\varepsilon = \frac{a^2}{\xi_1^2 \eta_1^2} + \frac{D_z \beta}{\xi_1 \eta_1 + \beta} \tag{39}$$

This relationship has no extremum with respect to parameter ξ_1.

In the theory of optimal filtration the magnitude $\varepsilon^* = \varepsilon - Z = X - Y$ is considered as an error. The second initial moment of this magnitude possesses extremal properties. Thus, under the conditions of the preceding example the magnitude α_ε^* is equal to:

$$\alpha_\varepsilon^* = \frac{a^2}{\xi_1^2 \eta_1^2} + \frac{D_z \eta_1 \xi_1}{\xi_1 \eta_1 + \beta} \tag{40}$$

This function has an extremum with respect to parameter ξ_1.

It is possible to measure directly the magnitude ε^* in tracking systems using *a priori* information on the statistical properties of the input useful signal and the disturbances. In practice it is possible to measure the error ε and the signal $Z_1 = Z_1(Z, X)$ related to Z. For instance, the function Z_1 may be obtained by filtering with special filters the input signal $X + Z$ and utilizing the information that the spectrum of the frequencies of the disturbance Z, as a rule, is substantially broader than the spectrum of the useful signal X. Then the function Z_1 will possess characteristics which are close to the characteristics of the function Z.

Measuring the magnitudes ε and Z_1 it is possible to formulate artificially a quality index having an extremal characteristic with respect to coefficient of amplification ξ_1 of the correcting circuit $B(\xi)$. For this the function of the error is assumed to have the form:

$$f(\varepsilon, \xi) = \varepsilon^2 + \psi(\xi_1) Z_1^2 \tag{41}$$

The function $\psi(\xi_1)$ may be chosen in a specific case, for instance, from the condition of proximity of the extrema of functions $M[\varepsilon - Z]^2$ and $M[\varepsilon^2 + \psi(\xi_1) Z^2)$ with respect to parameter ξ_1 for statistically prescribed input signal.

As an illustration of the method of prescribing a function $\psi(\xi_1)$ let us consider the case of good filtration when it is possible to neglect the component X in function Z_1. Let us determine $\psi(\xi_1) = \nu\xi_1$, where ν is a constant coefficient computed from the condition of proximity of the values of parameters ξ_{10} for extremal values of the functions $\tilde{\alpha} = M\varepsilon^2 + \nu\xi_1 D_z$ and $\alpha_\varepsilon^* = M(\varepsilon - Z)^2$.

In *Figure 3* there are presented graphs of functions $\tilde{\alpha}$ and α_ε^* corresponding to the minimal value and computed for the preceding example. For $\nu = 0{\cdot}1$ the minima of the functions (curves with an index 1) coincide closely, and the optimal value of parameter $\xi_{10} = 3{\cdot}0$. The change in a sufficiently broad range of probability characteristics of disturbance Z, useful signal X, and parameter η leads to a distortion of the form of the curves $\tilde{\alpha}$ and α_ε^*. However, their minima coincide, but are not reached for other values of parameter ξ_{10} as shown in *Figure 3*. In *Figure 3* the index 2 denotes curves for $D_2 = 10^{-3}$ and the previous values of other parameters.

In *Figure 4* there is shown a schematic diagram of a linear tracking system with tuning of the amplification coefficient ξ_1 for $\psi(\xi_1) = \nu\xi_1$. The function Z_1 is separated with the aid of a band pass filter or a filter of high frequencies. Then the signal is supplied to a square wave generator and a circuit with amplification coefficient $\nu\xi_1$, and then to a low frequency filter. Now consider the quasi-stationary regime of self-tuning of parameters. Eqn (31) of tuning of parameter ξ_1 stated with respect to deviation U_1 assumes the form:

$$[(TD+1)D - \lambda a_1] u_1 = -2\lambda m_{\varepsilon_0}[C_{10}(0) + C_{10}(D)]\varepsilon_0^0 - D\xi_0 + 2\lambda\nu m_z Z_1^0 \tag{42}$$

where

$$a_1 = M\{[C_{10}(D)\varepsilon_0^0]^2 + [\varepsilon_0 (C_{10}^2(D)\varepsilon_0^0]\} > 0 \tag{43}$$

From these one obtains the following equation for the determination of mathematical expectation m_{u1}:

$$(TD^2 + D - \lambda a_1) m_{u_1} = -D\xi_{10} \tag{44}$$

For $\lambda < 0$ the stable process of tuning is assured. When one determines the centred random component u_1^0, one obtains the equations:

$$[TD^2 + D - \lambda a_1] u_1^0 = -2\lambda m_{\varepsilon_0}[C_{10}(0) + C_{10}(D)]\varepsilon_0^0 + 2\lambda\nu m_{z_1} Z_1^0 \tag{45}$$

The magnitude m_{z_1} may be set equal to zero by proper selection of the corresponding filter. Taking this into account and also utilizing expressions for ε_0^0 in terms of $X^0 + Z^0$, one obtains from eqn (45)

$$u_1^0 = \Phi_1(D)(X^0 + Z^0) \tag{46}$$

where

$$\Phi_1(D) = \frac{-2\lambda m_{\varepsilon_0}[C_{10}(0) + C_{10}(D)]}{(TD^2 + D - \lambda a_1)[1 + A(D)B_0(D)]} \tag{47}$$

In this case, for computing the dispersion of parameter u_1 in a stabilized regime, one obtains:

$$D_{u_1} = \int_{-\infty}^{\infty} |\Phi_1(i\omega)|^2 [S_x(\omega) + S_z(\omega)]\, d\omega \tag{48}$$

where S_x and S_z are the spectral densities of random functions X and Z. For $\xi_{10} = \text{const.}$ the magnitude $m_{u_1} = 0$ in the stabilized regime. In this case the systematic error of a following system with self-tuning in a stabilized regime of operation is equal to $m_\varepsilon = m_{\varepsilon_0}$, that is, equal to systematic error for an optimal value of parameter ξ_{10}. The random component of the error of following is equal to:

$$\varepsilon^0 = \left[1 + \frac{b_1 A(D)}{1 + A(D)B_0(D)}\Phi_1(D)\right]\frac{1}{1 + A(D)B_0(D)}(X^0 + Z^0) \tag{49}$$

where the magnitude b_1 according to formula (36) is given by

$$b_1 = \frac{\partial B_0(\xi_0)}{\partial \xi_{10}} m_{\varepsilon_0} \tag{50}$$

In computing the dispersion of error ε one obtains the formula:

$$D_\varepsilon = \int_{-\infty}^{\infty} \left|\left[1 + \frac{b_1 A(i\omega)}{1 + A(i\omega)B(i\omega)}\Phi_1(i\omega)\right]\frac{1}{1 + A(i\omega)B(i\omega)}\right|^2 [S_x(\omega) + S_z(\omega)]\, d\omega \tag{51}$$

The calculations carried out for a tracking system (*Figure 4*) having the values of the preceding example for $\lambda = 10^5$, $T = 1{\cdot}0$, and the optimal value of parameter $\xi_{10} = 3{\cdot}0$, show a sufficiently good effectiveness of tuning. Thus, the mathematical expectation of tuned parameter ξ_1 is equal to $m_{\zeta_1} = \xi_{10}$, and the dispersion of the error of tuning computed by formula (48) is given by $D_{\zeta_1} = D_{u_1} = 4 \times 10^{-7}$. From these calculations it follows that the maximum relative error of tuning the parameter ξ_1, is equal to $6{\cdot}3 \times 10^{-2}$. per cent. As regards the error of tracking by the following system, the mathematical expectation of this error in tuning coincides with the value of this magnitude in an optimal system $m_\varepsilon = m_{\varepsilon_0} = 0{\cdot}33 \times 10^{-2}$.

The dispersion of the error of tracking in a self-tuning system computed by formula (51) coincides with a precision to three significant figures with a value of dispersion of the error of tracking in the optimal system $D_\varepsilon \approx D_{\varepsilon_0} = 2{\cdot}31 \times 10^{-5}$. Thus, in the considered example the self-tuning system with the utilization of the method of auxiliary operator assures an effective tuning for the minimum of the second initial moment of error in the presence of random disturbances.

Conclusion

The considered scheme of a self-tuning system may be effectively utilized both for the direct control of objects and the synthesis of automatic control systems during their design. The advantages of the system of self-tuning utilizing the method of auxiliary operator are: relative simplicity of achieving tuning circuits, effectiveness of operation in the presence of disturbances, and the possibility of obtaining high values of quick response.

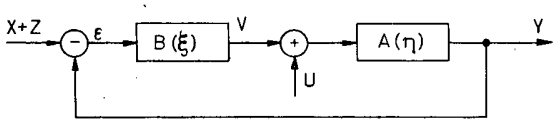
References

1 FELDBAUM, A. A. *Computers in Automatic Control Systems*. 1959. Moscow; GIFML

2 MARGOLIS, M. and LEONDES, C. T. A parameter tracking servo for control systems. *Trans. Inst. Radio Engrs, N. Y.* AC–4, N 2 (1959)

3 MARGOLIS, M. and LEONDES, C. T. On the theory of adaptive control systems; the learning model approach. *Automatic and Remote Control*. 1961. London; Butterworths

4 PUGACHYOV, V. S. Theory of random functions and its application to problems of automatic control. 1960. Moscow; GIFML



Figurg 1

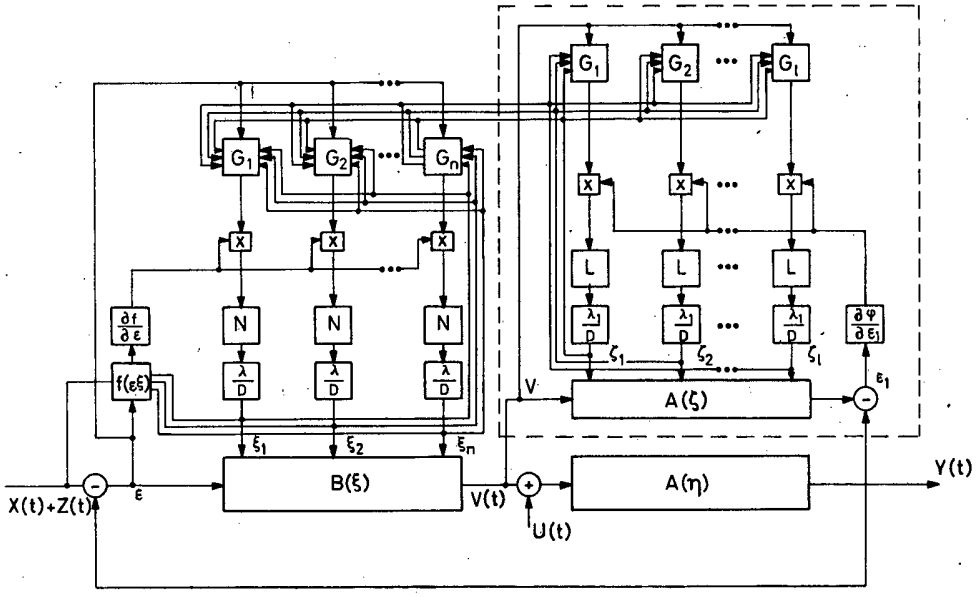


Figure 2

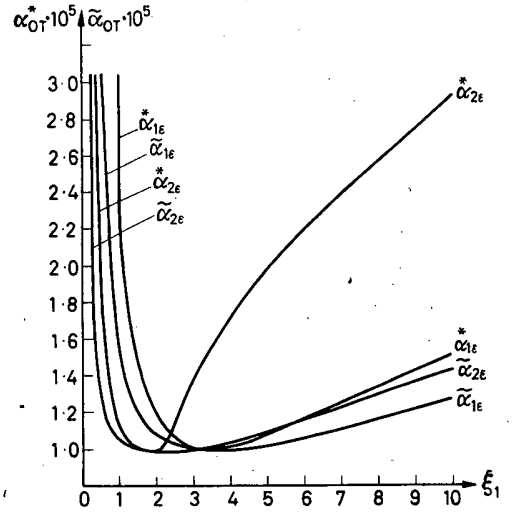


Figure 3

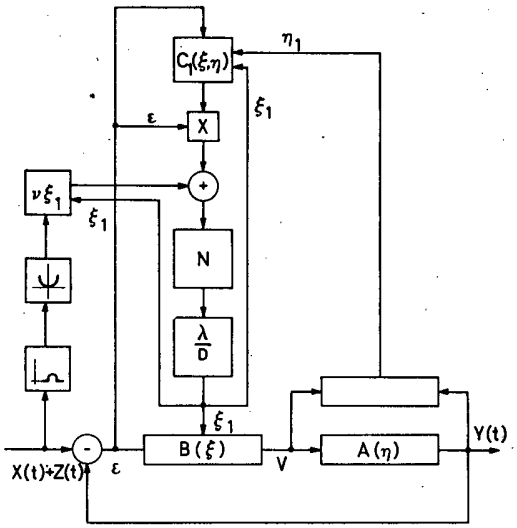


Figure 4

One Self-adjusting Control Systems Without Test Disturbance Signals

E.P. POPOV, G.M. LOSKUTOV and R.M. YUSUPOV

Statement of the Problem

In this paper, the term 'self-adjusting control system' means a system which performs the following three operations:

(1) Measures by means of automatic search or computes from the results of measurements the dynamic characteristics of the system, and possibly the characteristics of the disturbances as well.

(2) On the basis of this or that criterion defines the controller setting, parameters or structure needed for calibration (or optimization).

(3) Realizes the resultant controller structure, parameter or setting values.

Many studies of the theory and practice of self-adjusting control systems for stationary controlled plants have so far appeared in the world literature. There have also been contributions on self-adjusting of quasi-stationary systems. But there is almost a complete lack of contributions dealing more or less specifically with problems of synthesis and analysis of self-adjusting control systems for essentially non-stationary controlled plants. Moreover, as far as the authors are aware, even in the case of stationary and quasi-stationary systems, the process of self-adjustment is frequently effected solely on the basis of an analysis of the dynamic characteristics of the system, without taking into account the unmeasured external disturbances acting upon the controlled plant. At the same time it is obvious that external disturbance, besides the dynamic characteristics of the system, determines the quality of the process of control.

Another drawback of many of the self-adjusting systems in existence and proposed in the literature is the need to use special test signals to check the dynamic characteristics of the system.

This paper proposes, and attempts to validate, one of the possible principles for the creation of a self-adjusting control system for a particular class of non-stationary controlled plants.

The main advantage of the principle in question is the opportunity it provides to take account of both internal (system parameters) and external (harmful and controlling disturbances) conditions of operation of the system. In contrast to the self-adjusting systems known, a system created in accordance with the principle proposed will make it possible to obtain automatically the fullest possible information about the process under control without the use of test signals.

For the operation of a self-adjusting control system created on the basis of the principle proposed, a mathematical model of a reference (calculated) control system must be constructed. A 'reference system' is understood to be a system the controller of which is designed in accordance with the requirements on the quality of the control process, with the assumption that the mode of variation in time of the system's parameters as well as the disturbance effects is known.

The structure of the mathematical approximation of the real process is selected to match that of the mathematical model of the reference process. The self-adjusting system operates in such a way as to ensure continuous identity between the mathematical approximation of the real process and the model of the reference system. In this connection, the problem is posed of making the mathematical approximation of the real process as close as possible to the model of the reference process.

Without loss of generality, the case of control of only one variable is considered, which is denoted by x, and the corresponding reference differential equation is written in the form

$$x_E^{(n)} + \sum_{i=0}^{n-1} a_i^E(t)\, x_E^{(i)} = \sum_{i=0}^{m} b_m^E(t)\, f_E^{(i)} \qquad (1)$$

The real process is approximated by a linear differential equation of the same structure:

$$x^{(n)} + \sum_{i=0}^{n-1} a_i(t)\, x^{(i)} = \sum_{i=0}^{m} bm(t)\, f_E^{(i)} \qquad (2)$$

$$t = t_0,\ x^{(i)}(t_0) = x_{E0}^{(i)} \quad (i = 0, 1, \ldots, n-1)$$

The operation of the proposed self-adjusting control system will be examined in accordance with the sequence of the process of self-adjustment, indicated at the beginning of the definition.

General Case of Determination of the Dynamic Characteristics of a System

In order to create an engineering method of determining the dynamic characteristics of non-stationary systems in the construction of a self-adjusting control system, this paper proposes the use of the methods of stationary systems. For this purpose, the non-stationary system (1) is replaced by an equivalent system with piecewise-constant coefficients. (The methods of stationary systems are used on the intervals of constancy of the coefficients.) The transfer from a system with variable coefficients to one with piecewise-constant coefficients is effected on the basis of a theorem which can be formulated with the assistance of a number of the propositions of the theory of ordinary differential equations. In accordance with this theorem, the solution of a differential equation of form (1) with piecewise-continuous coefficients (a finite number of discontinuities of the first kind is assumed) can be obtained with any degree of accuracy in a preset finite interval (t_0, T_0) by breaking down the latter into a finite number of sub-intervals. (t_K, t_{K+1}) and replacement of

the variable coefficients within each sub-interval by constants, equal to any values of the corresponding coefficients inside or on the boundaries of the sub-intervals under consideration. In the general case, it is expedient to effect the breakdown process by the method of multiple iteration of solutions on a high-speed computer.

Let the differential equation with variable coefficients (1) be approximated by an equation with piecewise-constant coefficients.

Then, for $t \in (t_K, t_{K+1})$, one may write

$$x_E^{(n)} + \sum_{i=0}^{n-1} a_{iK}^E x_E^{(i)} = \sum_{i=0}^{m} b_{iK}^E f^{(i)} \tag{3}$$

In accordance with differential equation (3), the real process is approximated by the equation

$$x^{(n)} + \sum_{i=0}^{n-1} a_{iK} x^{(i)} = \sum_{i=0}^{m} b_{iK} f^{(i)} \tag{4}$$

As the dynamic characteristics of the system at the first stage of operation of the self-adjusting system on each interval (t_K, t_{K+1}), the coefficients a_{iK} $(i = 0, 1, \ldots, n-1)$, b_{iK} $(i = 0, 1, \ldots, m)$ are defined.

The simplest way to define these coefficients lies in defining the values of x and f and their corresponding derivatives at the points $t_K = \tau_1, \tau_2, \ldots, \tau_s = t_{K+1} - \Delta t$.

By substituting these values into eqn (4), one obtains for each interval (t_K, t_{K+1}) a system of S algebraic dissimilar equations for defining the searched coefficients.

In practice it is not always possible to measure the disturbing effect f and its derivatives. Therefore, in the general case, the above-mentioned method of defining the coefficients a_{iK} and b_{iK} cannot be directly employed.

This difficulty may be avoided in the following way. The real process is approximated, not by differential eqn (4), but by a differential equation of the form

$$\bar{x}^{(n)} + \sum_{i=0}^{n-1} \bar{a}_{iK} \bar{x}^{(i)} = \sum_{i=0}^{m} \bar{b}_{iK} f^{(i)} \tag{5}$$

In eqn (5) the disturbing effect and its corresponding derivatives are taken to equal the reference values. This avoids the need to measure the real disturbance f, and makes it possible to use the above-mentioned means of defining the coefficients of the differential equation approximating the real control process. The non-agreement of the real disturbances with the reference ones are taken into account through the coefficients a_{iK} and b_{iK}. Therefore dashes are placed over them.

In the general case $\bar{x}^{(i)} \not\equiv x^{(i)}$ $(i = 0, 1, \ldots, n)$ i.e., there is an approximation error. In view of this, in the transfer from eqn (4) to eqn (5), it is necessary to evaluate the maximum possible value of this approximation error, using for this purpose the assumed values of the limits of variation of disturbance f.

If for some class of controlled plants it can be assumed that in the process of operation only scale of the disturbance changes, i.e., the equality

$$f(t) = C_K f_E(t), \quad t \in (t_K, t_{K+1}) \tag{6}$$

where C_K is the random scale of disturbance, is satisfied, then the approximation error is absent, and the connection of the coefficients of eqns (4) and (5) is expressed by the equalities:

$$\begin{aligned} \bar{a}_{iK} &= a_{iK} \quad (i = 0, 1, \ldots, n-1) \\ \bar{b}_{iK} &= C_K b_{iK} \quad (i = 0, 1, \ldots, m) \end{aligned} \tag{7}$$

Equation (5) is used (henceforward, to simplify the notation, the dashes over the coefficients and the variable x are dropped) for definition of the coefficients a_{iK} and b_{iK}. It is assumed that measurements $x, x', \ldots, x^{(n)}$ are performed at the points $t_K = \tau_1, \tau_2, \ldots, \tau_S = t_{K+1} - \Delta t$.

The values of $f_E, f_E', \ldots, f_E^{(m)}$ are known. Then, for the definition of $(n + m + 1)$ desired coefficients in each interval (t_K, t_{K+1}) one obtains the following system of S algebraic equations, which will be written in abbreviated form thus:

$$\sum_{i=0}^{n-1} x^{(i)}(\tau_j) a_{iK} - \sum_{i=0}^{m} f_E^{(i)}(\tau_j) b_{iK} = -x^{(n)}(\tau_j) \quad (j = 1, 2, \ldots, S) \tag{8}$$

It is not always expedient to solve directly system (8) for $S = m + n + 1$, since, on account of the existence of measuring instrument errors and random high-frequency control process oscillations, the accuracy of definition of the coefficients will be very low. Moreover, for the same reasons, system (8) may be altogether incompatible.

To eliminate the case of incompatibility and to increase the accuracy of definition of the searched coefficients the method of least squares is employed[1, 2]. In so doing, the problem of approximation is also solved. When utilizing this method, it is expedient to take $S > m + n + 1$.

Using the method of least squares, the coefficients a_{iK}, b_{iK} are defined, minimizing according to these coefficients the function

$$L = \sum_{j=1}^{S} \rho(\tau_j) L_j^2$$

where

$$L_j = \sum_{i=0}^{n-1} x^{(i)}(\tau_j) a_{iK} - \sum_{i=0}^{m} f_E^{(i)}(\tau_j) b_{iK} + x^{(n)}(\tau_j)$$

is the disagreement, and $\rho(\tau_j)$ are weight coefficients which define the value of each measurement and, accordingly, of each of equation of system (8).

The necessary condition of the minimum of function L is the equality to zero of its first-order partial derivatives according to a_{iK} and b_{iK}. Having computed the partial derivatives and equated them to zero, one obtains an already compatible system of $m + n + 1$ linear algebraic equations for the definition of $m + n + 1$ coefficients:

$$\begin{aligned} \frac{\partial L}{\partial a_{iK}} &= \sum_{j=1}^{S} \rho(\tau_j) L_j \frac{\partial L_j}{\partial a_{iK}} = 0 \quad (i = 0, 1, \ldots, n-1) \\ \frac{\partial L}{\partial b_{iK}} &= \sum_{j=1}^{S} \rho(\tau_j) L_j \frac{\partial L_j}{\partial b_{iK}} = 0 \quad (i = 0, 1, \ldots, m) \end{aligned} \tag{9}$$

Solving system (9) by known methods, one obtains the values of a_{iK} and b_{iK}.

In certain cases the process of control at intervals may be approximated by a differential equation of the form

$$x^{(n)} + \sum_{i=0}^{n-1} a_{iK} x^{(i)} = \varphi_{EK}(t) \tag{10}$$

where

$$\varphi_{EK}(t)=\sum_{i=0}^{m} b_{iK}^{E} f_{E}^{(i)}(t)$$

This coarser approximation will make it possible to reduce computing time considerably by a reduction of the quantity of searched coefficients; in the given case only the coefficients a_{iK} are desired.

In the given approximation the deviations of the values of real coefficients b_{iK} and real disturbances f will be taken into account in the system *via* the values of the coefficients a_{iK}. System (11) will be the initial algebraic system for definition of the coefficients:

$$\sum_{i=0}^{n-1} x^{(i)}(\tau_j)\, a_{iK}=\varphi_{EK}(\tau_j)-x^{(n)}(\tau_j) \quad (j=1,2,\ldots,S) \qquad (11)$$

For definition of the searched coefficients a_{iK} by the method of least squares, one minimizes the function

$$L_1=\sum_{j=1}^{S} \rho(\tau_j)\, L_j^2 \qquad (12)$$

where

$$L_j=\sum_{i=0}^{n-1} x^{(i)}(\tau_j)\, a_{iK}+x^{(n)}(\tau_j)-\varphi_{EK}(\tau_j)$$

Using the necessary condition of the existence of a minimum of function (12) for the definition of n, coefficients a_{iK} ($i = 0, 1, \ldots, n-1$), one obtains a system of n algebraic equations:

$$\frac{\partial L_1}{\partial a_{iK}}=\sum_{j=1}^{S} \rho(\tau_j)\, L_j \frac{\partial L_j}{\partial a_{iK}}=0 \quad (i=0,1,\ldots,n-1) \qquad (13)$$

All the above discussion and the operations were performed on the assumption that the values of the control variable and the necessary quantity of derivatives at the moments of time of interest are available. In practice, however, one is usually limited to second-order derivatives.

In a number of cases real high-order systems may be approximated by second-order differential equations, preserving the description of their main dynamic properties. But even in the case of more complex high-order systems it is possible to suggest a number of algorithms for defining the searched coefficients, given the existence of a limited quantity of derivatives, some of which are as follows:

(*a*) Derivatives of higher orders of the control variable can be calculated with the assistance of a digital computer on the basis of the Lagrange and Newton interpolation formulae or according to the formulae of quadratic interpolation (method of least squares).

(*b*) If one integrates each term of eqns (5) and (10) $n - q$ times, where q is the order of the senior derivative of the control variable, which one can measure in a system with the requisite accuracy, then, taking the limits of integration t_K, τ_j ($j = 1, 2, \ldots, S$), one obtains the integral forms of eqns (8) and (11) respectively. If reference values are given to the magnitudes $x^{(n-1)}(t_K), x^{(n-2)}(t_K), \ldots, x^{(n-q+1)}(t_K)$ in these equations, then for defining the coefficients a_{iK} ($i = 0, 1, \ldots, n-1$) and b_{iK} ($i = 0, 1, \ldots, m$) it is sufficient to measure the derivatives to the qth order.

(*c*) Practically all existing controlled plants and control systems can be described by a set of differential equations, each of which characterizes one degree of freedom of movement and therefore has an order no higher than second.

(*d*) Sometimes, to reduce the order of the derivatives required for measurement, one may also take advantage of a number of coarse assumptions in relation to the terms of eqns (5) and (10), which contain derivatives of high orders.

For example, in these equations the values of the derivatives $x^{(n)}, x^{(n-1)}, x^{(n-q+1)}$ can be assumed equal to the reference values.

(*e*) The coefficients of approximating eqns (5) and (10) can be defined without any recourse to algebraic systems (8) and (11), if one uses the following method[5].

Let the composition of the control system include an analogue simulator, on which is set up a differential equation of form (5) or (10). In this simulator there is a controlling device, which provides an opportunity to effect variation of coefficients a_{iK} and b_{iK} in a certain way.

The control system memorizes the curve of the real process in the interval $(t_K, t_{K+1} - \Delta t)$, and selection of the coefficients a_{iK} and b_{iK} is performed on the simulator in such a way as to bring together in a certain sense the real process and the solution of the equation set up on the simulator.

When the quantitative value of the proximity evaluation reaches the predetermined value, the magnitudes of coefficients a_{iK} and b_{iK}, are fixed and extracted for subsequent employment in the self-adjusting control system. Obviously the simulator operation time scale must be many times less than the real time scale of the system. Only under this condition can the requisite high speed of self-adjustment be achieved. Practically any time scale may be realized with the assistance of analogue computing techniques.

Automatic Synthesis of Controller Parameters

For the operation of the majority of self-adjusting systems, the system operation quality criterion is set in advance. For systems constructed on the basis of the proposed principle, it is generally expedient to use as the criterion the expression

$$M=\sum_{i=0}^{n-1} (a_{iK}-a_{iK}^{E})^2+\sum_{i=0}^{m} (b_{iK}-b_{iK}^{E})^2 \qquad (14)$$

This criterion generalizes both the methods of approximation of the real control process expounded above.

To simplify subsequent operations, the following notations are introduced.

$$b_{0K}=a_{nK}; \quad b_{1K}=a_{n+1,K}, \ldots, b_{mK}=a_{m+n,K}$$

Expression (14) can then be rewritten in the form

$$M=\sum_{i=0}^{n_0} (a_{iK}-a_{iK}^{E})^2; \qquad n_0=\begin{cases} n+m \text{ for (5)} \\ n-1 \text{ for (10)} \end{cases} \qquad (15)$$

On each interval (t_K, t_{K+1}) the adjustable parameters are so selected as to bring expression (15) to the minimum. The ideal, i.e., most favourable, case would be one when M would reach zero as the result of selection of the adjustable parameters. This is not always possible, however. In the first place, not all the coefficients a_{iK} ($i = 0, 1, \ldots, n_0$) are controllable. Second, in multi-loop non-autonomous systems even the values of the controllable coefficients cannot all be tuned up to the reference values simultaneously, since the relationship of the coefficients

a_i to the adjustable parameters, although usually linear, is nevertheless arbitrary in relation to the quantity of adjustable parameters, the sign and the coefficients with which these parameters enter into expressions for a_i.

The second difficulty may be avoided by means of successful selection of the reference system or by complete disconnection of the loops (channels) of control of the main variables, i.e., by satisfying the conditions of autonomy.

It is assumed that all the coefficients a_i ($i = 0, 1, \ldots, n_0$) are controllable (in practice the values of uncontrollable coefficients may be reckoned to be reference values). Then, for the coefficients a_i one may write

$$a_i = a_i(K_1, K_2, \ldots, K_p; T_1, T_2, \ldots, T_q; l_1, l_2, \ldots, l_r)$$
$$(i = 0, 1, \ldots, n_0)$$

where $K_1, K_2, \ldots, K_p$ are the gains of the controlled plant; $T_1, T_2, \ldots, T_q$ are the time constants of the controlled plant and the controller, and $l_1, l_2, \ldots, l_r$ are the gains of the controller (adjustable parameters).

Since the coefficients a_i usually depend on the adjustable parameters linearly, one may write

$$a_i = \sum_{j=1}^{r} \mu_{ij} l_j + \nu_i \quad (i = 0, 1, \ldots, n_0) \tag{16}$$

where

$$\mu_{ij} = \mu_{ij}(K_1, K_2, \ldots, K_p; \quad T_1, T_2, \ldots, T_q);$$

$$\nu_i = \nu_i(K_1, \ldots, K_p; T_1, T_2, \ldots, T_q)$$

Using the necessary condition for the existence of a minimum of function M, one obtains the following algebraic system for determination of the setting values $l_1, l_2, \ldots, l_r$

$$\sum_{i=0}^{n_0} [a_{iK}(l_1, l_2, \ldots, l_r) - a_{iK}^E] \frac{\partial a_{iK}(l_1, l_2, \ldots, l_r)}{\partial l_j} = 0$$
$$(j = 1, 2, \ldots, S) \tag{17}$$

It is assumed that when the system is in operation, the adjustable parameter values only change in accordance with their computed values, i.e., at any moment of time one knows the magnitudes of $l_1, l_2, \ldots, l_r$. Then, for the interval (t_K, t_{K-1}) until the moment of correction of the adjustable parameters in accordance with expression (16), one can write:

$$a_{iK} = \sum_{j=1}^{r} \mu_{ijK} l_{j, K-1} + \nu_{iK} \tag{18}$$

From system (18) one may determine the magnitudes of M_{ijK} and ν_{iK} ($i = 0, 1, \ldots, n_0$; $j = 1, 2, \ldots, r$) since the values of a_{iK} ($i = 0, 1, \ldots, n_0$) and $l_{j, K-1}$ ($j = 1, 2, \ldots, r$) are known.

Taking into account eqn (16), after substitution of the values of M_{ijK} and ν_{iK} the algebraic system (17) for defining $l_{1K}, l_{2K}, \ldots, l_{rK}$ takes the form

$$\sum_{i=0}^{n_0} \left[\left(\sum_{j=1}^{r} \mu_{ijK} l_{jK} + \nu_{iK} \right) - a_{iK}^E \right] \mu_{ijK} = 0 \tag{19}$$
$$(j = 1, 2, \ldots, r)$$

Realization of Adjustable Parameters

Block-circuit with a Self-adjusting System using a Digital Computer

The duration of the intervals of constancy of the coefficients of reference eqn (3), when a digital computer is used in the control system, must satisfy correlation

$$t_{K+1} - t_K = T_1 + T_2 + T_3 + \Delta t \tag{20}$$

where $T_1 = \Delta\tau\,(S - 1)$ is the time required to carry out measurements; $T_2 = N/n_0$ is the time required for the computations; T_3 is the time of actuator generation; $0 \leq \Delta t \leq t_{K+1} - t_K$; $\Delta\tau = \tau_{j+1} - \tau_j$ is the period of measurements ($j = 1, 2, \ldots, S$); n_0 is the computer speed of action, and N is the number of operations required to define coefficients l_{jK} ($j = 1, 2, \ldots, r$).

It is obvious that to ensure better operation of the self-adjusting system, it is necessary to reduce as much as possible the magnitude $T = T_1 + T_2 + T_3$.

Now the opportunities for reducing the time T_3 are dealt with. This question is directly linked with the choice of the actuator. Electromechanical servosystems with a considerable time constant are usually employed as actuators at the present time. But it turns out that it is possible to suggest a number of purely circuit variants of the change of the transfer functions or of gains of the correcting devices (regulators) of the system. These inertia-less actuators are termed 'static'. It is particularly advantageous to produce static actuators with the aid of non-linear resistors (varistors), valves with variable gains (varimu), electronic multipliers, etc.

Consider, for example, one of the variants of a static actuator based on an electronic multiplier. Let the made of control have the form

$$y = \sum_{j=1}^{r} l_j x^{(j)}$$

and let the jth adjustable parameter have the value l_j^0 at moment t_0^{j-1} (start of operation of the system). While the system operates in accordance with the signals of the computer, the value l_j is constantly being corrected.

Thus, at the end of the interval (t_K, t_{K+1}) one has

$$l_{jK} = l_j^0 + \Delta l_{jK}$$
$$y = \sum_{j=1}^{r} l_j^0 x^{(j)} + \sum_{j=1}^{r} \Delta l_{jK} x^{(j)} \tag{21}$$

Obviously each addend in the right-hand side of expression (21) can be instrumented with the aid of the circuit in *Figure 1*, where *EM* is the electronic multiplier, and *AD* the adder.

The following are self-adjusting system computer operating algorithms: when the real process is approximated by differential eqns (5), the algebraic systems (9), (18), and (19); when the real process is approximated by differential eqns (5), the algebraic systems (13), (18), and (19).

It is obvious that in the general case it is more convenient to solve the problem of self-adjustment according to the proposed principle with the aid of a high-speed digital computer. It can be specialized for solving systems of algebraic equations. *Figure 2* shows the block diagram of a self-adjusting system with a digital computer.

Some Particular Cases

In the preceding sections the proposed principle for creating a self-adjusting control system for non-stationary objects was expounded in general form. In practice, one may naturally encounter cases when the given principle can be used in more simplified variants. Several such opportunities are considered.

(1) Obviously, the entire theory expounded above can be applied fully to stationary and quasi-stationary systems, which are particular instances of non-stationary systems. In this case the durations of the intervals of constancy of the coefficients (t_K, t_{K+1}) equal, for stationary systems

$$K=0, t_{K+1}-t_K=t_1-t_0=T_0-t_0 \qquad (22)$$

for quasi-stationary systems

$$t_{K+1}-t_K \geq \Delta t_\rho \qquad (23)$$

where $\Delta t\rho$ is the control time (duration of the transient process).

As can be seen from relations (22) and (23), in stationary and quasi-stationary systems one is less rigidly confined to the time of analysis of the real process and synthesis of controller parameters. It is therefore possible to define coefficients a_{iK} and b_{iK} more accurately and to use criteria which reduce the self-adjustment process speed, but make it possible to increase the accuracy of operation of the system. Among such criteria one may cite, in particular, the integral criteria for the evaluation of the quality of a transient process[3].

For stationary and quasi-stationary systems the problem of self-adjustment in accordance with the principle proposed above may be solved as a problem of the change in position of the roots of the transfer function of a closed system, i.e., the self-adjustment problem may be solved in accordance with the requirements of the root-locus method, which is extensively employed in automatic control theory. A feature of the use of the proposition of the root-locus method in accordance with the principle under consideration is that the zeros and poles defined by the coefficients a_{iK} and b_{iK} are fictions since they not only depend on the parameters of the controlled plant and controller, but also depend on real disturbances as well.

(2) In practice, one may encounter cases when a controller is required to ensure only the stability of a system in the course of operation. As is known, the stability of linear stationary systems is determined by the coefficients of the characteristic equation. This proposition is also valid for certain quasi-stationary systems (method of frozen coefficients).

Therefore to solve the problem posed (the provision of stability), the control system must define the actual values of the coefficients of the left-hand side of the differential equation of the system and must set on the controller such gains factors as will satisfy the conditions of stability, for example the conditions of the Hurwitzian algebraic criterion. On the assumption that disturbance f is constant in the interval (t_K, t_{K+1}) the coefficients of the characteristic equation of the system on this interval are determined in the following way.

The differential equation of the system for $t \in (t_K, t_{K+1})$ is written in the form

$$x^{(n)}+\sum_{i=0}^{n-1} a_{iK}\, x^{(i)}=F_K$$

where F_K is in the general case the unknown right-hand side, constant for $t \in (t_K, t_{K+1})$. The algebraic system for determining the described coefficients will then be written thus:

$$x^{(n)}(\tau_j)+\sum_{i=0}^{n-1} x^{(i)}(\tau_j)\, a_{iK}=F_K \quad (j=1,2,\ldots,S) \qquad (24)$$

Since F_K is unknown, but is constant in the interval (t_K, t_{K+1}) it is eliminated with the assistance of one of the equations of system (24). For this purpose one uses the equation

$$x^{(n)}(\tau_l)+\sum_{i=0}^{n-1} x^{(i)}(\tau_l)\, a_{iK}=F_K \quad (1 \leq l \leq S)$$

After eliminating F_K one has:

$$\sum_{i=0}^{n-1} [x^{(i)}(\tau_j)-x^{(i)}(\tau_l)]\, a_{iK}=-[x^{(n)}(\tau_j)-x^{(n)}(\tau_l)]$$
$$(j=1,2,\ldots,l-1,l+1,\ldots,S) \qquad (25)$$

By resolving system (25) directly with $S=n+1$, or by the least-squares method with $S>n+1$, one determines the coefficients a_{iK} $(i=0, 1, \ldots, n-1)$, the values of which are used, if the need arises, for synthesis of the values of the controller parameters which ensure the stability of the system.

Conclusion

The paper has expounded only the basis of the proposed principle for the construction of a self-adjusting control system in general form and in certain particular cases. Studies are under way on problems connected with the approximation of differential equations with essentially variable coefficients by differential equations with piecewise-constant coefficients, with the selection of the type of computer to operate in the self-adjustment loop, with the dynamic precision of the self-adjustin system, etc. The investigations which have been made allow one to hope that the use of the principle expounded in this paper for the construction of self-adjusting control systems will prove extremely effective in many cases when it is expedient to use the natural oscillations of the system, without introducing test disturbance signals.

References

1 GONCHAROV, V. L. *The Theory of Interpolation and Approximation of Functions.* 1954. Moscow; Gostekhizdat

2 LANDOSH, K. *Practical Methods of Applied Analysis.* 1961. Moscow; Fizmatgiz

3 POPOV, E. P. *Dynamics of Automatic Control Systems.* 1954. Moscow; Gostekhizdat

4 CHERNETSKII, V. I., and YUSUPOV, R. M. On one type of adaptive control system. *Izv. Akad. Nauk SSSR, Otdel Tekhn. Nauk* (1962)

5 VITENBERG, I. M. Specialized electric analog with automated scan. *Use of Computer Techniques for Production Automation.* 1961. Moscow; Mashgiz

6 POPOV, E. P. *Automatic Regulation and Control.* 1962. Moscow; Fizmatgiz

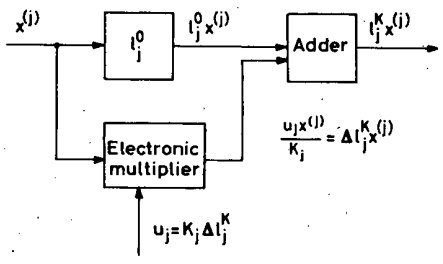


Figure 1

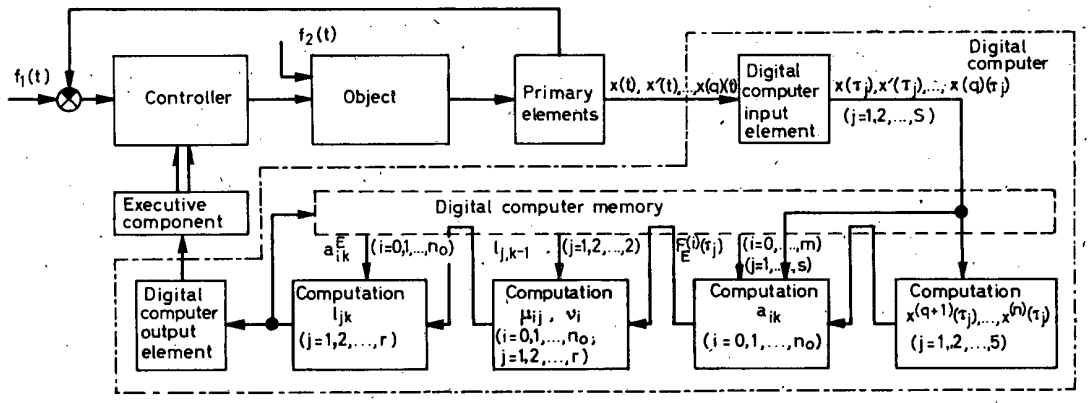


Figure 2

Optimal Processes in Systems with Time Lag

N.N. KRASOVSKII

Introduction

The problem of forming the optimal process input for a regulator in a system with time lag of action and signals is considered in this paper. The questions considered belong to the class of problems of optimal control. These problems were first stated and developed in the U.S.S.R by Feldbaum[1]. The mathematical theory of optimal processes was worked out by Pontryagin *et al.*[2], on the basis of their Maximum Principle. Their studies have given rise to a great number of works: for instance, that by Rozonoer[3], and also the *Theory of Dynamic Programming*[4], developed by Bellman and his colleagues on the basis of the optimality principle and the functional equations which follow from it, which embraces a very wide class of problem. Reference can be made to the authors whose works, among others, have a direkt connection with this paper[5–18].

Reference can also be made to the works of those authors who, among others, have studied optimal control problems in after-action systems, and in more general systems with distributed parameters[19–22].

The present work originates from the studies of Letov[23, 24], and the statement of the problem adopted here is a generalization, for systems with after-action, of the statement of the problem given by Letov[24]. The problems for systems with delay of the feedback signals considered below are related to problems of dual control[25] or of the theory of adaptive processes[26].

The solution proposed is based on the method of Liapunov functions and the theory of stability of motion[27, 28], developed for equations with time lags[29], and modernized in accordance with the principles of *Dynamic Programming*[4]. Statements of the problems are given in this paper, and criteria of optimality and the principles of solution are formulated. For systems which can be described by a few actual equations, the explicit analytical form of the optimal regulator is given. Approximate methods for calculating optimal control are described, and problems complicated by random circumstances considered.

Time-lag of Signals in the Plant

Consider a controlled system (*Figure 1*) where $z(t)$ is a controlled vector quantity at the output of the plant A, and ξ, a scalar quantity, is the input of the regulator B, constituted on the basis of information on the actual error $x = z - z^0$, and possibly also on the actual values of the load $\eta(t)$. The special feature of the system is the time-lag of the signals in the plant A (Case I), or of signals in the feedback channels (1) and (2) (Case II), or of ξ in channel (3) (Case III). Each case will be examined separately. If Cases I—III are combined in one system, the statement of the problems and the solutions must be combined accordingly.

Case I. Assume that the disturbed motion of the system is described by the equation

$$\frac{dx}{dt} = f[t, x(t), x(t-h_1), \ldots, x(t-h_k), \eta(t), \xi] \qquad (1)$$

where x is an n-dimensional error vector, h_i is the time lag of signals in the plant ($0 < h_i \leq h$, $i = 1, \ldots, k$), f is a known vector function of its own arguments, determined by the structure of the system, and $\eta(t)$ is the load or disturbance. Besides this, a functional determining the quality of the process is given, and there may be a restriction on the magnitude of the control signal ξ.

The disturbed motion $x(t)$ of system (1) with after-action, with $t > t_0 \geq 0$ is determined, as is well known, by the history $x(t_0 + \theta)$ $(-h \leq \theta \leq 0)$ of this motion. The initial function $x(t_0 + \theta)$ $(-h \leq \theta \leq 0)$ will therefore be called the initial disturbances (with $t = t_0$). It is also convenient to consider, as quantities describing the state of system (1) at instants $t \geq t_0$, and determining its future motion when $\tau > t$, sections of the trajectories $x(t + \theta)$ $(-h \leq \theta \leq 0)$. It is therefore suitable to form the control signal $\xi(t)$ at each instant t on the basis of information on the whole of the realized trajectory $x(t + \theta)$ with $-h \leq \theta \leq 0$. In other words, analytic construction of the regulator[24] means finding ξ in the form of a some functional $\xi(t) = \xi[t, x(t + \theta)]$, determined on the curves $x(t + \theta) = \{x_i(t + \theta), -h \leq \theta \leq 0, i = 1, \ldots, n\}$. In future it will be assumed that the argument θ varies within the limits $-h \leq \theta \leq 0$. The continuous functions $x(\theta)$ or $x(t + \theta)$ of the argument θ are assumed to be elements of a certain space X with a matrix

$$\|x(\theta)\| = \max_{\theta} (x_1^2(\theta) + \ldots + x_n^2(\theta))^{\frac{1}{2}}$$

Also used is the notation

$$\|x(0)\| = (x_1^2(0) + \ldots + x_n^2(0))^{\frac{1}{2}},$$

$$\|x(t)\| = (x_1^2(t) + \ldots + x_n^2(t))^{\frac{1}{2}}$$

Three problems are considered:

Problem 1. Find a control signal $\xi = \xi^0\, t, x(\theta)$ such that the motion $x = 0$ in a closed system (1) (that is, with $\xi(t) = \xi^0(t, x(t + \theta))$ is asymptotically stable[29] with respect to the disturbances $x^0(t_0 + \theta)$ $(t_0 \geq 0)$ from a region

$$\|x^0(\theta)\| \leq G_0 \qquad (2)$$

and such that for all $t_0 \geq 0$ and $x^0(t_0 + \theta)$ out of (2) there holds a minimum

$$J[t_0, x^0, \xi^0] = \min_{\xi} J[t_0, x^0, \xi] \qquad (3)$$

Here

$$J[t_0, x^0, \xi] = \int_{t_0}^{\infty} \omega[t, x(t, t_0, x^0, \xi), \xi(t)]\, dt \qquad (4)$$

where ω is a given non-negative function, $x(t, t_0, x^0, \xi)$ is the trajectory of (1) with initial conditions t_0 and $x^0(t_0 + \theta)$ and a selected law of control $\xi(t) = \xi[t, x, (t + \theta)]$. The control signal ξ can be constrained by a supplementary restriction $\xi \in \Xi$ (for instance, $|\xi| \leq 1$).

Problem 2. Find a control signal $\xi = \xi^0 t, x(\theta)$ assuring a minimum of

$$J_T[t_0, x^0, \xi^0] = \min_{\xi \in \Xi} J_T[t_0, x^0, \xi] \quad (0 \leq t_0 \leq T) \tag{5}$$

where

$$J_T[t_0, x^0, \xi] = \int_{t_0}^{T} \omega[t, t_0, x^0, \xi), \xi(t)]\,dt + \psi[x(T, t_0, x^0, \xi)] \tag{6}$$

and $T < \infty$ is a given instant of time, while $\|x^0(t_0 + \theta)\| \leq G_0$.

Problem 3. Find a control signal $\xi = \xi^0[t, x(\theta)]$ assuring minimum of

$$J_\infty[t_0, x^0, \xi^0] = \min_{\xi \in \Xi} J_\infty[t_0, x^0, \xi] \tag{7}$$

where $\|x_0(t_0 + \theta)\| \leq G_0$ and

$$J_\infty[t_0, x^0, \xi] = \lim \frac{J_T}{T - t_0} \text{ when } T \to \infty \tag{8}$$

In Problems 2 and 3, as in 1, it is assumed that the initial conditions x^0 and trajectories $x(t, t_0, x^0, \xi^0)$ do not go beyond certain previously fixed regions.

The sufficient conditions of optimality of the control signal ξ^0 will be formulated for Problems 1 and 2.

Theorem 1. Let it be possible to indicate functionals $v[t, x(\theta)]$ and $\xi^0[t, x(\theta)]$, defined and satisfying in some region $\|x(\theta)\| \leq G$ the following conditions:

(1) The functional v is positive definite with respect to $\|x(0)\|$.

(2) The functional v admits an upper limit with respect to $\|x(\theta)\|$.

(3) The following inequality is satisfied:

$$\text{in } f[v[t, x(\theta)] \text{ when } \|x(0)\| = G,$$

$$\|x(\theta) = G] \geq \sup[v[t, x(\theta)] \text{ when } \|x(\theta)\| \leq G_0]$$

(4) Along trajectories of (1)[29] the derivative $(dv/dt)_\xi$ of the functional v satisfies the condition

$$\left(\frac{dv}{dt}\right)_{\xi_0} + \omega[t, x(t), \xi^0] = \min_{\xi \in \Xi}\left[\left(\frac{dv}{dt}\right)_{\xi} + \omega[t, x(t), \xi]\right] = 0 \tag{9}$$

in the region $\|x(t + \theta)\| \leq G$, and is negative definite with respect to $\|x(t)\|$ in this region.

Then $\xi^0[t, x(t + \theta)]$ is the optimal control signal for Problem 1, and the following equality is valid:

$$v[t_0, x^0(t_0 + \theta)] = J[t_0, x^0(t_0 + \theta), \xi^0] \tag{10}$$

Note. Properties (1) and (2) generalize in a natural way the corresponding properties of Liapunov's functions[27] that is (1) means that there exists a function $w(r) > 0$ with $r \neq 0$, such that $v[t, x(\theta)] \geq w(\|x(0)\|)$ with $\|x(\theta)\| = \|x(0)\|$, and (2) means that there exists a function $W(r)$ satisfying the conditions $W(0) = 0$, $v[t, x(\theta)] \leq W(\|x(\theta)\|)$. If in Problem 1 the region G_0 encompasses any possible large initial disturbances x_0 (the problem of optimal stabilization as a whole), the region G must coincide with the whole of the space X, and (1) is replaced by the condition

$$\lim v[t, x(\theta)] = \infty \text{ when } \|x(0)\| \to \infty, \|x(\theta) = \|x(0)\| \tag{11}$$

uniformly with respect to t.

The demonstration of Theorem 1 is made by reasoning typical for the theory of stability of motion[29], but taking into account the principles of dynamic programming[4].

The sufficient criterion of optimality for Problem 2 is formulated as follows:

Theorem 2. Let there exist for every $\|x^0(t_0 + \theta) \leq G_0$ and $t_0 \in [0, T)$ an admissible control signal $\xi(t)$, that is, a control signal for which the trajectory $x(t, t_0, x^0, \xi)$ may be prolonged in some finite region G until the instant $t = T$, and therefore the integral (6) is finite. If one can find in the region G functionals $v[t, x(\theta)]$ and $\xi^0[t, x(\theta)]$ satisfying conditions (9), and

$$v[T, x(\theta)] = \psi[x(0)] \tag{12}$$

then ξ^0 is the optimal control signal for Problem 2, and the following equality is valid:

$$v[t_0, x^0(t_0 + \theta)] = J_T[t_0, x^0(t_0 + \theta), \xi^0] \tag{13}$$

The solution of Problem 3 can be obtained by passage to the limit from the solution of the problem when $T \to \infty$.

Note. If the load (t) is random or the system is subject to random disturbance, Problems 1 to 3 are modified as follows: integrals (4), (6) and (8) are replaced by their mathematical expectations (the conditional mathematical expectations for the appropriate initial conditions t_0, x^0, η^0), and in Problem 1 the requirement of stability is replaced by the requirement of stochastic stability[30]. In this case seek the control signal ξ^0 in the form of a functional $\xi^0[t, x(t + \theta), \eta(t + \tau)]$, where $-h \leq \theta \leq 0$ and $-h^* \leq \tau \leq 0$, while $h^* = 0$ is the value of the maximal after-action for the probability process $\eta(t)$ (if $\eta(t)$ is a Markov process, then $h^* = 0$). The criteria of optimality given above preserve their form, with the modification that v must here also be a functional $v[t, x(\theta), \eta(\tau)]$, and the derivative $(dv/dt)_\xi$ is replaced by its average value[30] $(dM\{v\}/dt)_\xi$.

Conditions (9) reduce to partial derivative equations of a special kind. The solution of these equations in the general case is cumbersome; it is possible, however, to indicate a number of cases when an explicit form can be found for the optimal control signal, or when a numerical procedure for its determination can be indicated.

The results of applying the proposed criteria to systems described by equations of actual form will be illustrated.

Let the transient process be described by the linear differential equations

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} a_{ij}(t)\,x_j(t) + \sum_{j=1}^{n} c_{ij}(t)\,x_j(t - h) + b_i\xi + a_i\eta(t) \tag{14}$$

where a_{ij}, c_{ij}, a_i and b_i are known functions of time or constants. First assume that $\eta(t) \equiv 0$, and then consider Problem 1 for

system (14), assuming that

$$J=\int_{t_0}^{\infty}\left[\sum_{i=1}^{n}x_i^2(t)+\lambda\xi^2(t)\right]\mathrm{d}t,\quad \lambda>0-\text{const} \qquad (15)$$

any initial disturbances $x^0(t_0+\theta)$ are admissible.

Here the functional v from Theorem 1 must be chosen in the form

$$v[t,x(\theta)]=\sum_{i,j=1}^{n}\Big[\mathrm{d}_{ij}(t)\,x_i(0)\,x_j(0)+2\,x_i(0)\int_{-h}^{0}\beta_{ij}(t,\theta)\,x_j(\theta)\,\mathrm{d}\theta+\int_{-h}^{0}\int_{-h}^{0}\gamma_{ij}(t,\theta,\tau)\,x_i(\theta)\,x_j(\tau)\,\mathrm{d}\theta\mathrm{d}\tau\Big] \qquad (16)$$

which generalizes in a natural way the Liapunov function widely used in stability theory, as a quadratic form. If for every initial condition x^0, t_0 there exists an admissible control signal $\xi(t)$, that is, a control signal (t) for which integral (15) converges uniformly with respect to t_0, then there exists a functional v (16) satisfying the conditions of Theorem 1. From this it is directly concluded that in this case there exists an optimal control signal ξ^0 having the form

$$\xi^0[t,x(t+\vartheta)]=\sum_{i=1}^{n}\left[\mu_i(t)\,x_i(t)+\int_{-h}^{0}v_i(t,\vartheta)\,x_i(t+\vartheta)\,\mathrm{d}\vartheta\right] \qquad (17)$$

Conclusion

The optimal regulator ξ^0 in system (14) with condition of minimum (15) is seen to be the regulator B, which applies to the input of the controlled plant A at every instant t a quantity ξ^0 (17), worked out on the basis of a measurement of the error x at the given instant of time t and at previous instants $t-h\leq\tau\leq t$, while the results of measurement of the previous errors $x(\tau)=x(t+\vartheta)$ must be processed in the integrators $\int v_i(t,\vartheta)\,x_j(t+\vartheta)\,\mathrm{d}\vartheta$. The control signal ξ^0 depends linearly on $x(t+\vartheta)$ $(-h\leq\vartheta\leq 0)$.

It is interesting to observe that for a system (14) with discrete delay $h>0$ the optimal control signal must be worked out by an element with continuous distribution of the after-action $v_i(t,\vartheta)$ over the whole of the time-lag interval $-h\leq\vartheta\leq 0$.

Now let $\eta(t)\not\equiv 0$ be a known function of time. Consider for system (14) the problem (2), where

$$J_T=\int_{t_0}^{T}\left[\sum_{i=1}^{n}x_i^2(t)+\lambda\xi^2(t)\right]\mathrm{d}t+\sum_{i,j=1}^{n}\psi_{ij}\,x_i(T)\,x_j(T) \qquad (18)$$

Here any restricted control signal $\xi(t)$ is admissible, and the following assertion is valid: a functional v satisfying the conditions of Theorem 2 exists, and differs in form from the functional (16) by the term

$$v^*=\sum_{i=1}^{n}(\delta_i(t)\,x_i(0)+\int_{-h}^{0}\varphi_i(t,\vartheta)\,x_i(\vartheta)\,\mathrm{d}\vartheta)+\varrho(t) \qquad (19)$$

From this assertion follows the conclusion that in this case an optimal control signal always exists, and differs in form from the control signal (17) by a term $\xi^*=\varkappa(t)$ which is a function only of time t.

Note. The conditions for Problem 1 Solvability for systems (14) and (15) reduce to the possibility of constructing an admissible control signal $\xi(t)$. Here, as also in the case of systems without delay, the question is connected with the conditions of controllability of the system[11, 31]. System (14) (with $\eta(t)\equiv 0$) will be called fully controlled in the interval $[t_0, t_1]$ $(t_1>t_0+h)$ provided that for every initial condition $x^0(t_0+\theta)$ there exists a continuous (piece-wise-continuous) control signal $\xi(t)$ such that $x(t, t_0, x^0, \xi)\equiv 0$ when $t_1-h\leq\leq t\leq t_1$. The conditions of controllability, as in the case without delay[31], can be investigated starting from the 'L problem'. If system (14) is fully controllable in every sufficiently long section of the t axis, then it is optimally stabilizable in the sense of Problem 1. It is also observed that such stabilization is certainly possible if system (14) is asymptotically stable with $\xi=0$, or if the delay $h>0$ is sufficiently small (or if the c_{ij} are small), and for the system $\mathrm{d}x_i/\mathrm{d}t=\Sigma a_{ij}\,x_j+b_i\xi$ the conditions of full controllability are fulfilled: the vectors $\{b_i\}$, $\{\|a_{ij}\|\{b_i\}\}$, ..., $\{\|a_{ij}\|^{n-1}\{b_i\}\}$ are linearly independent. The conditions of solvability of Problem 1 for (14) and (15) can also be ascertained in the process of solution, if the solution is sought by passage to the limit from the solution of Problem 2 for (14) and (18) (with $\psi_{ij}=0$, $\eta=0$ and with $T\to\infty$), which is sometimes a convenient method in practice.

Now consider Problem 3 for system (14): accept that in (8) $\omega=\sum_{i=1}^{n}x_i^2+\lambda\,\xi^2$ and assume $\eta(t)$ to be a random Markhov function (for definiteness, of the pure discontinuous or diffusion type). Moreover, assume that system (14) is subject to some irregular disturbance of the white noise type, causing diffusion spread of $x(t)$ in the time $\mathrm{d}t$ with a matrix of second moments $\|M\{\mathrm{d}x_i\,\mathrm{d}x_j\}\|=\|\sigma_{ij}(t)\,\mathrm{d}t\|$.

The following result is obtained: if system (14) with $\eta(t)\equiv 0$ is stabilized in the sense of Problem 1, then an optimal control signal ξ^0 exists and has the form

$$\xi^0[t,x(t+\theta),\eta(t)]=\sum_{i=1}^{n}\left[\mu_i(t)\,x_i(t)+\int_{-h}^{0}v_i(t,\theta)\,x_j(t+\theta)\,d\theta\right]+\kappa(t,\eta(t)) \qquad (20)$$

It is interesting to observe that the first term here tallies with (17), and the random term $\kappa(t,\eta(t))$ determined by the actual values of $\eta(t)$ is the same as it would be if, with $\tau>t$, the function $\eta(\tau)$ were determined and tallied with the prediction of its mathematical expectation $M\{\eta(\tau)/\eta(t)\}$ made according to the actual value of $\eta(t)$. The magnitude of the dispersion of $\eta(t)$ and the quantities $\sigma_{ij}(t)$ do not affect ξ^0, and manifest themselves only naturally in the quantity $M\{J_\infty[t_0, x^0, \eta^0, \xi^0]\}$.

As has been shown above, it is very laborious, in the general case, to construct the functional from Theorems 1 and 2. The following methods may be indicated for approximate its determination (and consequently that of ξ^0): the small parameter method; approximate solution of the functional equation (9); approximating v in the mean; replacing the equations with delays or the functional equation (9) by finite difference equations; replacing the equation with delays by a set of equations

for the Fourier coefficients of a section of the trajectory $x(t+\theta) - h \leq \theta \leq 0$. These methods can be illustrated by numerical examples.

Delay of Feedback Signals

Consider now the system of *Figure 1* when there is no after-action in the plant A, but signals in channels 1—3 can be delayed.

Case II. Let the motion of the plant A be described by the vector differential equation

$$\frac{dx}{dt} = f[t, x(t), \eta(t), \xi] + \dot{\phi} \tag{21}$$

where x, η, ξ, f have the same meaning as in the first part of the paper, and $\dot{\phi}$ is a disturbance of the white noise type, giving rise to diffusion spread of $x(t)$ in the time dt with the matrix

$$\|M\{dx_i\, dx_j\}\|_1^n = \|\sigma_{ij}(t)\|_1^n\, dt \tag{22}$$

The problem is to minimize the quantities

$$J_T = M\left\{\int_{t_0}^{T} \omega[t, x(t), \xi(t)]\, dt + \psi[x(T)]\right\} dt \tag{23}$$

and

$$J_\infty = \lim \frac{J_T}{T - t_0} \text{ with } T \to \infty \tag{24}$$

The peculiarity of the case in question is that information concerning the actual values of the error $x(t)$ and load $\eta(t)$ are supplied by way of channels 1 and 2 with delays of $h_1 > 0$ and $h_2 > 0$ (or either $h_1 > 0$ or $h_2 > 0$) respectively ($h_1 \leq h$, $h_2 \leq h$). In other words, assume that in the regulator B at the instant t in the closed interval $[0, T]$ the values of the actual quantities $x(t - h_1)$ and $\eta(t - h_2)$, where $\eta(t)$ is a random Markov function, are known. Also assume that the regulator B is capable of remembering up to the instant t the signal $\xi(t+\theta)$ worked out by it with $-h \leq \theta < 0$. Denote the set of magnitudes $x(t - h_1)$, $\eta(t - h_2)$ and $\xi(t + \theta)$ $(-h \leq \theta < 0)$ by $y(t)$, and $x(-h_1)$, $\eta(-h_2)$, $\xi(\theta)$ $(-h \leq \theta < 0)$ by respectively y. The quantity $y(t)$ makes it possible to compose a probability description of the plant A at the instant t. The quantities J_T (23) and J_∞ (24) with the chosen law of control ξ may be regarded as functionals with respect to $y(t_0)$, that is,

$$M\left\{\int_{t_0}^{T} \omega[t, x(t), \xi(t)]\, dt + \psi(T)\right\} = J_T[t_0, y^0(t_0), \xi] \tag{25}$$

$$\lim_{T \to \infty} \frac{J_T}{T - t_0} = J_\infty[t_0, y^0(t_0), \xi] \tag{26}$$

It is therefore reasonable in this case to seek the optimal control signal ξ^0 as a function of $y(t)$, that is, in the form of a functional

$$\xi(t) = \xi[t, y(t)] \tag{27}$$

Call the admissible control signals the set of such functionals, sufficiently regular to give a meaning to the solution of (21) with $\xi(t)$ of (27), and, possibly, constrained by supplementary restrictions arising from the statement of the problem (for instance, $|\xi| \leq 1$). Designate the set of admissible control signals by the symbol Ξ. Now the problem can be formulated.

Problem 4. It is required to find a control signal ξ^0 belonging to Ξ which minimizes (25) for all y^0 belonging to Y_0, $t_0 \geq 0$.

Problem 5. It is required to find a control signal ξ^0 belonging to Ξ minimizing (26) for all y^0 belonging to Y_0, $t_0 \geq 0$. Here Y_0 is some region of the components y given in advance.

Denote by $x(t, y^0(t_0), \xi)$ the random motion of the system, generated by the initial conditions $y^0(t^0)$ with a certain choice of the control law; moreover, assume necessarily, with $t_0 = h \leq \leq t < t_0$, that the control signal $\xi(t)$ tallies with that $\xi(t_0 + \theta)$ $(t_0 + \theta = t)$ which is a component of $y^0(t_0)$.

Now formulate the criterion of optimality for Problem 4.

Theorem 3. It is assumed that for all $y^0(t_0)$ belonging to Y_0 and $0 \leq t_0 \leq T$ there exists an admissible control signal $\xi(t)$ (or $\xi = \xi[t, y(t)]$) such that (25) has a meaning, is finite, and almost all the realizations $\{x(t, y^0(t_0), \xi),\ \eta(t, y^0(t_0)),\ \xi(t+\theta)$ $(-h \leq \theta < 0)\}$ belong to Y, where Y is a certain region of values of y. Let it be possible to find functionals $v[t, y]$ and $\xi^0[t, y]$ satisfying the conditions

$$(1) \qquad v[T, y(T)] = M\{\psi[x(T, y(T), \xi)]\} \tag{28}$$

for all $y(T)$ belonging to Y

$$(2) \qquad \left(\frac{dM\{v\}}{dt}\right)_{\xi^0} + M\{\omega[t, x(t, y(t), \xi^0), \xi^0]\}$$

$$= \min_{\xi \in \Xi}\left[\left(\frac{dM}{dt}\right) + M\{\omega[t, x(t, y(t), \xi), \xi]\}\right] = 0 \tag{29}$$

for all $y(t)$ belonging to Y and all t in the closed interval $[0, T]$.

Then $\xi^0[t, y(t)]$ is the optimal control signal for Problem 4 and $v[t_0, q^0(t_0)] = \min J_T[t_0, y^0(t_0), \xi]$.

The solution of Problem 5 is obtained by passage to the limit from the solution of Problem 4.

The results of applying the given criterion to a system described by equations of an actual form are illustrated. Consider Problem 5 for the system

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} a_{ij}(t)\, x_j(t) + b_i \xi + a_i \eta(t) + \dot{\phi} \tag{30}$$

with the condition of minimum (26), where

$$J_T = M\left\{\int_{t_0}^{T}\left[\sum_{i,j=1}^{n} \omega_{ij}(t)\, x_i(t)\, x_j(t) + \lambda \xi^2(t)\right]\right\} dt$$

$$+ \sum_{i,j=1}^{n} \psi_{ij}\, x_i(T)\, x_j(T) \tag{31}$$

The delays along both channels 1 and 2 are assumed to be equal to $h > 0$, and it is admitted that any initial deviations $x^0(t_0 - h)$ and $\eta(t_0 - h)$ belong to (η_1, η_2).

With sufficiently wide assumptions concerning the character of the Markov probability process $\eta(t)$ and with the condition of full controllability of the system $dx_i/dt = \Sigma a_{ij}\, x_j + b_i\, \xi$, the functionals $v[t, y]$ and $\xi^0[t, y]$ satisfying criterion (21) can be found, and passage to the limit with $T \to \infty$ can be carried out. Problems 4 and 5 can also be solved. In addition the following result is valid.

Results

The optimal control signal for Problems 4 and 5 stated with conditions (30) and (31) has the form

$$\xi^0[t, y(t)] = \sum_{i=1}^{n} \mu_i(t)\, x_i(t-h) + v[t, \eta(t-h)] + \int_{-h}^{0} \varrho[t, \theta]\, \xi[t+\theta]\, d\theta \qquad (32)$$

The term v is determined at every instant t with respect to the realized $\eta(t-h)$, but to calculate it one must know the prediction $M\{\eta(\tau)/\eta(t-h)\}$ with $\tau > t-h$.

Here the functional $v[t, y(t)]$ has the form of the sum of the quadratic and linear functionals of $x_i(t-h)$ and $\xi(t+\theta)$ with coefficients dependent on $\eta(t-h)$.

Analysing the resulting solution ξ^0 the following conclusion is arrived at: the optimal control signal ξ^0 chosen here at every instant t is the same as would be obtained in a deterministic system and without delay of the feedback signals; however here, instead of the known quantities $x_i(t)$ of the deterministic system, their best mean square predictions $M\{x_i(t)/x(t-h), \eta(t-h), \xi(t+\theta)(-h \leq \theta < 0)$ must enter into the control law, and the deterministic load $\eta(\tau)$ $(\tau > t-h)$ is likewise replaced by the mean prediction $M\{\eta(\tau)/\eta(t-h)\}$.

Case III. This case reduces naturally to the previous one, and it is not considered individually.

When several of the cases analysed are combined in one system, the statements of the problems, criteria of optimality and results are combined correspondingly.

In conclusion it is observed that Case II can be included in the more general case when incomplete information is transmitted the feedback channels 1 and 2. For it can be assumed along indeed that at the instant t there are applied to the regulator B signals $y(t)$ and $\zeta(t)$, statistically connected with $x(t)$ and $\eta(t)$ (in Case II, $\{y(t), \zeta(t)\} = \{x(t-h), \eta(t-h), \xi(t+\theta)\}$) and an optimal control signal in dependence of these signals can be constructed. The foregoing reasoning and conclusions are generalized to this more general case. The quality of the process depends on how much the processes $\{y(t), \zeta(t)\}$ and $\{x(t), \eta(t)\}$ are connected informationally, or, in other words, how far the processes $\{x(t), \eta(t)\}$ are observable[12] with respect to $\{y(t), \zeta(t)\}$.

References

[1] FELDBAUM, A. A. Optimal processes in automatic control systems. *Automat. Telemech.* 14, No. 6 (1953)

[2] PONTRYAGIN, L. S., BOLTYANSKII, V. G., GAMKRELIDZE, R. V., and MISHCHENKO, E. F. A mathematical theory of optimal processes. Fizmatgiz (1961)

[3] ROZONOER, L. I. Pontryagin's Maximum Principle in the theory of optimal systems. *Automat. Telemech.* 20, Nos. 10–12 (1959)

[4] BELLMAN, R. *Dynamic Programming. I.I.L.* (1960)

[5] LERNER, A. YA. Maximum high speed of automatic control systems. *Automat. Telemech.* 15, No. 6 (1956)

[6] LERNER, A. YA. *Design Principles for High-speed Following Systems and Regulators.* 1961. Moscow; Gosenergoizdat

[7] BELLMAN, R., GLICKSBERG, J., and GROSS, O. *Some Aspects of the Mathematical Theory of Control Processes.* 1958. Project Rand

[8] FELDBAUM, A. A. Calculating devices in automatic systems. *Fizmatgiz* (1959)

[9] FILIPPOV, A. F. Some questions of the theory of optimal control. *Vestn. MGU* No. 2 (1959)

[10] KALMAN, R. E., and BERTRAM, J. E. Control systems analysis and design via the 'Second Method' of Liapunov. *Pap. Amer. Soc. Mech. Eng.*, No. 2 (1959)

[11] KALMAN, R. E. On the general theory of control systems. *Automatic and Remote Control.* 1961, Vol. 2. London; Butterworths

[12] KALMAN, R. E. New methods and results in linear prediction and filtering theory. *RJAS Tech. Rep.* 61–1 (1961)

[13] KULIKOVSKI, R. A. *Bull. Acad. Polon. Sci., Serie des sciences techniques*, Vol. VII, No.. 6, 11, 12 (1959), Vol. VIII, No. 4 (1960)

[14] LA SALLE, J. Time optimal control systems. *Proc. nat. Acad. Sci.*, Vol. 45, No. 4 (1959)

[15] GIRSANOV, I. V. Minimax problems in the theory of diffusion processes. *Dokl. AN SSSR*, Vol. 136, No. 4 (1960)

[16] TSYPKIN, YA. Z. On optimal processes in pulsed automatic systems. *Dokl. AN SSSR*, Vol. 136, No. 2 (1960)

[17] BELLMAN, R., and KALABA, R. Theory of dynamic programming and control systems with feedback. *Automatic and Remote Control.* 1961, Vol. 1. London; Butterworths

[18] MERRIAM, K. U. Calculations connected with one class of optimal control systems. *Automatic and Remote Control.* 1961, Vol. 3. London; Butterworths

[19] BUTKOVSKII, A. G., and LERNER, A. YA. Optimal control of systems with distributed parameters. *Dokl. AN SSSR*, Vol. 134, No. 4 (1960)

[20] KRAMER, J. On control of linear systems with time lags. *Inform. Control*, Vol. 3, No. 4 (1960)

[21] KHARATISHVILI, G. L. The maximum principle in the theory of optimal processes with time lags. *Dokl. AN SSSR*, Vol. 136, No. 1 (1961)

[22] BELLMAN, R., and KALABA, R. Dynamic programming and control processes. *J. Bas. Engng.* (March 1961)

[23] LETOV, A. M. Analytical construction of regulators. *Automat. Telemech.*, Vol. 21, Nos. 4–6 (1960)

[24] LETOV, A. M. Analytic construction of regulators; the dynamic programming method. *Automat. Telemech.*, Vol. 22, No. 4 (1961)

[25] FELDBAUM, A. A. Information storage in closed systems of automatic control. *Izv. AN SSSR, Otdelenie tekhnicheskikh nauk. Energet-automat.*, No. 4 (1961)

[26] BELLMAN, R. Adaptive control processes. 1961. *Project Rand*

[27] LIAPUNOV, A. M. *The General Theory of Stability of Motion.* 1950. Gostekhizdat

[28] CHETAEV, N. G. *Stability of Motion.* 1956. Gostekhizdat

[29] KRASOVSKII, N. N. *Some Problems of the Theory of Stability of Motion.* 1960. Gostekhizdat

[30] KRASOVSKII, N. N., and LIDSKII, E. A. Analytic construction of regulators in systems with random parameters. *Automat. Telemech.* Vol. 22, Nos. 9–11 (1961)

[31] KRASOVSKII, N. N. A problem of tracking. *Prikl. matemat. mech.*, Vol. 26, No. 2 (1962)

[32] KRASOVSKII, N. N. Analytic construction of an optimal regulator in a system with time lags. *Prikl. matemat. mach.* Vol. 26, No. 1 (1962)

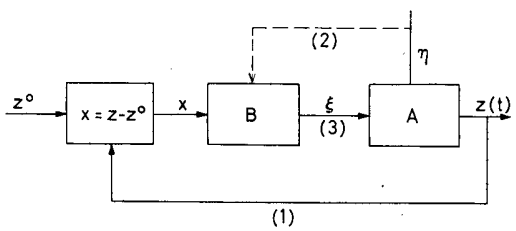


Figure 1.

Problems of Continuous Systems Theory of Extreme Control of Industrial Processes

A. A. KRASOVSKI

Many continuous industrial processes lend themselves to the following plan. There is available some quantity n of adjustments or controls of machines, apparatus, regulators securing an industrial process. The flow of the industrial process and the parameters depend on the coordinates of the adjusting or control elements (adjustment parameters).

Together with the controlling adjusting element coordinates the output parameters are affected by various disturbance factors (change of material parameters, wear of machines and tools, temperature and moisture variations and other factors).

The output parameters are controlled continuously or discretely (but with sufficiently small intervals of discontinuity by special measuring devices—output parameter information transmitters (*Figure 1*) influenced by disturbance factors and also by random variations of adjustment parameters. The output parameters are subjected to continuous variations.

Even though a practically ideal adjustment of the machine system, securing the industrial process, is initially attained, after some time the disturbance factors will bring forth considerable changes in the output parameters. In order to prevent the dropping out of the output parameters from the established tolerances (scrap output), adjustment and tuning of the machine system is necessary. Various means of automation of these operations are possible. If it is precisely known which parameter and to what extent it is affected by one or another controlling adjusting element, the usual feedback principle may be used (regulation by defiation). For this it is necessary first to smooth the results of measurements in order to eliminate overshoots of the system in the presence of small, random deviations within the tolerance limits. Methods of such automatic processing of information may be set up, based on the widely utilized methods of non-automated statistical control[2]. The measured and smoothed signals of output parameter deviations are conveyed to the performing arrangements and cause changes in the controlling adjusting element coordinates. Such systems are sometimes called staistical autotmata[3].

Undoubtedly the introduction of statistical automata will prove to be an important step in the automation of industry. However, a necessary condition of their application must be a sufficiently complete *a priori* information about the characteristics of the industrial process. In many cases this information is absent, and even if it is available during the initial period of the systems adjustment it loses authenticity in time, due to the change in properties of the industrial process.

Under these conditions the application of usual, non-self-adjusting control loops (statistical automata) becomes impossible. In these cases it is expedient to utilize an extremal control.

The present work is devoted to the investigation of some possible schemes of extremal control systems with continuous industrial processes and some questions of the theory of these systems. It is a development of earlier work by the author[1].

For the realization of an extremal control a quality output (production) index Q is selected, having extrema at wanted values of product parameters. Such an index may be, for example, the sum of the squares of deviation of the output parameters from the standard values. The quality index Q is determined by a computer (calculating machine in diagram *Figure 1*) based on information transmitter data on current values of output parameters. To secure the basic function of the system-maintenance of the quality index at the extremum level, search oscillations are necessary. Natural high frequency random oscillations, as well as artificially produced oscillations of controlling elements, may be employed as search oscillations. Naturally the first method is preferable, since it is not linked with any increase of high frequency fluctuations of the production parameters.

In order to make use of natural oscillations as search oscillations, it is necessary to measure them. The measurement of search oscillations is done by information transmitters for these oscillations (*Figure 1*), which measure controlling element oscillations and disturbance effects transmitted to them.

The measured search oscilations are transmitted to a simulator or a dynamic model of the industrial process. The purpose of the simulator is to transform the search oscillations in the same manner as these oscillations are transformed in a real process. For many industrial processes the simulator may be carried out in the shape of a delay line.

The output signals of the simulator are transmitted to the multiplying elements, to the other entrances of which is transmitted the computer signal which is proportional to the current value of the production quality index. The output values of the multiplying element are smoothed by the low frequency filters and are transmitted to the entrances of the control devices which move the controlling element.

If the quality index deviates from the extremum value, then a correlated component of the search fluctuations appears at the computer outlet. Values of the mathematical expectation of the duplicating links signals differing from zero then appear. Slowly changing signals are separated out by the low frequency filters and start the control devices. The controlling elements act on the production parameters in the direction approaching the extremum of the quality index.

The values of control parameters, together with the disturbance effects transmitted to them, are designated as X_ν ($\nu = 1, 2, \ldots, n$). Each control parameter brought forth has three components.

$$X_v = X_v^* + \delta X_v + \delta X_{vw}$$

Here X_v* working elements are output values of the extremum controlling portion of the system; δX_v are search elements for which it is expedient to utilize high frequency controlled effects transmitted to the control parameters. and δX_{vw} are uncontrolled disturbance effects transmitted to the control parameters.

The current value of the production quality index in general is a function of indicated control parameters and disturbance effects $f_1, f_2, \ldots, f_m$ according to transmitted control parameters.

When the transient process characteristics are described sufficiently accurately by time delays, then the current value of the production index is expressed by the function of preceding values of indicated control parameters and disturbances effects.

$$Q = Q[X_1(t-\tau_1), \ldots, X_n(t-\tau_n), f_1, \ldots, f_m] \qquad (1)$$

The selection of the composition of control parameters must conform to the following condition. To each set of permanent control parameter values must correspond a definite (with an accuracy up to the level of noises) set of production parameter values. In other words, in a static regime and with absence of noises a unilateral conversion of control parameters into production parameters must be realized. It should be noted that no mutual unilateral conversion is required, so that the number of control parameters may greatly exceed the number of controlled production parameters.

In virtue of one-sided-unilateral conversion, to each extremal function of production parameters, corresponds an extremal function of control parameters.

As agreed, the production quality index is an extremal function of its parameters. Therefore, function (1) in relation to the control parameters $X_1, X_2, \ldots, X_n$ is also extremal.

Adjustment-loss Time

Assuming that a process having unchanged, fixed working components of control parameters X_v*, is under investigation, and assuming also that, by the initial adjustment, it was possible, at some time $t = t_0$, to attain the extremum value of the production quality index, then under the influence of distrubance factors the production quality index will in time deteriorate spontaneously, in spite of the constancy of the control coordinates (*Figure 2*). At the expiration of time T_1 the quality index will get out of the permissible limits. The disturbance effects are random functions of time or random values, although in some individual applications their mathematical expectations may dominate centred random elements.

The chance of producing quality index with time $Q(t)$ is also a random time function, known to be non-stationary for this process with a fixed adjustment. And so, repeating the above test, one gets new realizations $Q(t)$ and new time values T_i (*Figure 2*).

The overall adjustment-loss time by the quality index Q is designated as the mathematical expectation $M(T_i)$ of time intervals T_i. So the overall adjustment-loss time expresses the mean value of time interval, after which the production quality index of the industrial process with a fixed adjustment gets out of the permissible limits.

The adjustment-loss time, understandably, depends on the nature of the industrial process and its automation level by means of frequency automatic systems. If the overall adjustment-loss time is great, then a non-automatic, hand control is not difficult and there is no need to use a complex self-adjusting system. If the aggregate adjustment-loss time is small, then a person is unable to secure adjustment even with the presence of appropriate data transmitters and self-adjustment becomes necessary.

It should be noted that the higher the speed of the industrial process and the stricter the demands on the quality of production, the smaller is the overall adjustment-loss time. Acceleration of the industrial processes and stepping up of demands on the quality of production are inherent characteristics of technical progress. Therefore, the application of self-adjusting control systems of industrial processes has a broad prospect.

Equations of Extreme Control Processes

It is assumed that, in the vicinity of the extremum point, serving as a working portion of the system under consideration, the quality index (1) approximates with sufficient accuracy by the quadratic shape of preceding values of cootrol parameters and by the additional member δQ_f expressing the influence of disturbing effects $f_1, \ldots, f_m$:

$$Q(t) = Q_l + \frac{1}{2}\sum_{i,j=1}^{n} a_{ik}\Delta k_i(t-\tau_i)\Delta X_j(t-\tau_j) + \delta Q_f \; a_{ij} = a_{ji} = \frac{\partial^2 Q}{\partial X_i \partial X_j} \qquad (2)$$

here

$$\Delta X_v = X_v - X_{vl} = X_v^* + \delta X_v + \delta X_{vw} - X_{vl} = \Delta X_v + \delta X_v + \delta X_{vw}$$

are complete deviations of brought forth coordinates (parameters) of control , $\Delta X_v^* = X_v^* - X_{vl}$ are working deviations of control coordinates, and Q_l is the extremum value of the quality index. In case the computer of the quality index does not bring about smoothing (smoothing is secured only by subsequent elements of the circuits) and the production parameter measuring instruments are practically non-inertial, or their inertness is accounted for in the values of time delays τ_i, the output value of the computer equals:

$$U(t) = Q(t) + \delta Q_n(t)$$

Here $\delta Q_n(t)$ is the element created by the errors of the production parameter meters and the errors of the computer. Thus

$$U(t) = Q_l + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}\Delta X_i(t-\tau_i)\Delta X_j(t-\tau_j) + \delta Q \qquad (3)$$

where

$$\delta Q = \delta Q_f + \delta Q_n$$

The value $U(t)$ in the multiplying elements of the synchronous detectors (correlators) is multiplied by the search signals δX_v displaced in time in the delay simulator. The errors in delay simulation are designated $\delta\tau_v$.

To the second entrances of the multiplying elements are

transmitted values $\delta X_1 (t - \tau_v - \delta\tau_v)$ and the output signals of these elements equal $V_v = U(t)\,\delta X_v (t - \tau_v - \delta\tau_v)$.

The linear portion of the controlling system without any common restriction is divided into a set of filters and integrating elements (*Figure 1*). The output working coordinates equal

$$X_k^* = \frac{1}{D} \sum_{v=1}^{n} W_{kv}(D)\, V_v$$

Here $\| W_{kv}(D) \|$ is the matrix of the transfer functions at low frequencies. Thus

$$DX_k^* = D\Delta X_k^* + D X_{kl} = \sum_{v=1}^{n} W_{kv}(D)\, V \quad D = \frac{d}{dt}$$

or

$$D\Delta X_k^* = \sum_{v=1}^{n} W_{kv}(D) [u(t)\,\delta X_v (t-\tau_v-\delta\tau_v)] - DX_{kl}$$

utilizing expressions (3) for $U(t)$, one finds

$$\begin{aligned} D\Delta X_k^* = {} & \frac{1}{2} \sum_{i_2,\, j_1,\, v} a_{ij} W_{kv}(D) \{[\Delta X_i^* (t-\tau_i) + \delta X_i (t-\tau_i) \\ & + \delta X_{iw}(t-\tau_i)] \times [\Delta X_j^* (t-\tau_j) + \delta X_v (t-\tau_j) + \delta X_{jw}(t-\tau_j)] \\ & \times \delta X_1 (t-\tau_v-\delta\tau_v)] \\ & + \sum_{v} W_{kv}(D)\, [(Q_l + \delta Q)\,\delta X_v (t-\tau) - \delta\tau_v)] - DX_{kl} \qquad (4) \\ & (k = 1, 2, \ldots, n) \end{aligned}$$

Summation by indices i, j, v is carried out within the limits from 1 to n.

Qualitative Analysis of Extremum Control Processes

Quasi-stationary Regime

The quality demand of an extremum control process reduces to the following. With considerable initial deviations from extremum the state point must move to the extremum as smoothly as possible (without much overshoot). In a steady operation the state point must stay sufficiently close to the extremum.

Let eqn (4) be converted into:

$$\begin{aligned} D\Delta X_k^* = {} & \sum_{i,\, j,\, v} a_{ij}\, W_{kv}(D) \\ & [\Delta X_i^* (t-\tau_i)\,\delta X_j (t-\tau_j)\,\delta X_v (t-\tau_v-\delta\tau_v)] \\ & + \frac{1}{2} \sum_{v} W_{kv}(D) \\ & \sum_{i,\, j} a_{ij} \Delta X_i^* (t-\tau_i)\,\Delta X_j^* (t-\tau_j)\,\delta X_v (t-\tau_v-\delta\tau_v) \\ & + \sum_{i,\, j,\, v} a_{ij}\, W_{kv}(D) \\ & [\Delta X_i^* (t-\tau_i)\,\delta X_{jw}(t-\tau_j)\,\delta X_v (t-\tau_v-\delta\tau_v)] \\ & + \delta\phi_k - DX_{kl} \qquad (5) \end{aligned}$$

here

$$\begin{aligned} \delta\phi_k = {} & \frac{1}{2} \sum_{i,\, j,\, v} a_{ij}\, W_{kv}(D) \{[\delta X_{iw}(t-\tau_i) + \delta X_i (t-\tau_i)] \\ & \times [\delta X_{jw}(t-\tau_j) + \delta X_j (t-\tau_j)]\,\delta X_v (t-\tau_v-\delta\tau_v)\} \\ & + \sum_{v} W_{kj}(D)\, [(Q_E + \delta Q)\,\delta X_v (t-\tau_v-\delta\tau_v)] \qquad (6) \end{aligned}$$

Values $\delta\phi_k$ may be treated as the effect of errors, noises and search elements, brought to the outputs of filters of the synchronous detectors, provided there are no working deviations ($\Delta X_i^* = 0$). These functions do not depend on working deviations (it is assumed, that δQ does not depend on working deviations) and on the whole may only obstruct the movement of the state point to the extremum.

Thus, $\delta\varphi_k$ always plays the role of disturbance effects and it is expedient to decrease them as much as possible. If the search elements δX_j, have permanent constituents then, as seen from expression (6), it is impossible to decrease indefinitely $\delta\varphi_k$ by any increase of time constant filters of the synchronous detectors. Indeed, according to (6), the constant components δX_v will cause deviations at the outputs of the synchronous detectors.

$$\frac{1}{2} \sum_{i,\, j,\, v} a_{ij}\, W_{kv}(0)\, \overline{\delta X}_i\, \overline{\delta X}_j\, \overline{\delta X}_v + \sum_{v} W_{kv}(0)\, (Q_v + \overline{\delta Q})\, \delta X_v$$

Where at least part of the transfer coefficients $W_{kv}(0)$ is known to differ from zero, since otherwise the circuit of the extremum control is inefficient. Thus, it is expedient to secure zero parity of the permanent elemeents of search constituents i.e. the centering of the search oscillations. This is easily attained by installation of high frequency filters at the outputs of the search oscillation pickups.

In particular, an ideal high frequency filter separates, from the input value, the high frequency constituent not correlated with the remaining part of the input value. This is illustrated by the graphs in *Figure 3*, showing a density spectrum curve $S(\omega)$ of the input function, which is assumed to be stationary and ergodic and amplitude frequency characteristic $A(\omega)$ of an ideal high frequency filter.

An ideal filter separates the high frequency constituent with a spectral density $S\delta_v(\omega)$ *Figure 3(b)* not correlated with the filtered component (spectral density $Sw(\omega)$), since the mutual spectral density of these components equals zero.

If the data meter controls the full input coordinate of the system $X_v = X_v^* + \delta X_v + \delta X_{vw}$, then the ideal high frequency filter in a stabilized operation separates the high frequency constituent δX_v not correlated with constituent $X_v^* + \delta X_{vw}$. It should be noted that stationary X_v^* may be expected only in a stabilized regime of the system operation. In transient regimes X_v^* is a non-stationary random function and even with the use of ideal filters the search elements prove to be to some extent correlated with the working elements X_v^*.

However, as is seen from the following in the present system (perhaps even more than in other continuous extremum systems), a quasi-stationary regime is profitable. In a quasi-stationary regime the transient process times are great compared to correlated times of search elements. When a quasi-stationary condition is secured and with the application of high frequency

filters near to ideal the search elements may be considered with a high degree of accuracy not correlated with X_v^*, both in a stabilized and in a transient condition of the system.

Based on the above the search elements δX_v it is assumed as centred by random functions not correlated to ΔX^*, δX_w, δQ. Investigation of other members of the right portions of eqn (5) is now made. The second member of the right portion may be rewritten in the shape

$$\frac{1}{2}\sum_{v} W_{kv}(D)\left[F^*\delta X_v(t-\tau_v-\delta\tau_v)\right]$$

where

$$F^*=\sum_{i,j}\Delta X_i^*(t-\tau_i)\,\Delta X_j^*(t-\tau_j) \tag{7}$$

In view of the definiteness of the signs of the functions of working deviations this member cannot facilitate the organization of movement to the extremum.

Thus, members (7) play the part of impeding effects and it is expedient to reduce their influence to the minimum values.

The only accepted means of reducing the effects of these members is the raising of frequencies (decreasing the correlation times) of the search elements at given times of transient processes of a closed loop or, inversely, increasing cumulative times at given correlation times of search elements. Either one or the other means switch to a quasi-stationary regime. In a quasi-stationary regime the effects of members (7) can be neglected. The following members of eqns (5)

$$\sum_{i,j,v} a_{ij}\,W_{kv}(D)\left[\Delta X_i^*(t-\tau_i)\,\delta X_{jw}(t-\tau_j)\,\delta X_v(t-\tau_v-\delta\tau_v)\right] \tag{8}$$

although linearly depend on working deviations, are also playing the role of impeding effects.

In fact, as agreed δX_{jw} and δX_v are not correlated and δX_v are centred. Therefore, the mathematical expectations of the products $\delta X_{jw}(t-\tau_j)\,\delta X_v(t-\tau_v-\delta\tau_v)$ equal zero. Thus, the expressions in the square brackets represent linear forms of working deviations, whose coefficients are centred 'high frequency' random time functions. These members can only increase the scattering of the trajectories of the state point during its movement to the extremum.

In the quasi-stationary regime, because of the intensive suppression of the high frequency constituents, members (8) may be neglected.

Turning to the investigation of the first member of the right portions of eqns (5) it is noticed that the product of search constituents may be represented as a sum of the mutual correlated (at $j\neq v$) or a auto-correlated function and a centred random function. Moreover, if the search constituents are stationary and are stationary combined, then the correlation functions depend only on the argument difference.

$$\delta X_j(t-\tau_j)\,\delta X_v(t-\tau_v-\delta\tau_v)=R_{jv}(\tau_v-\tau_j+\delta\tau_v)+\xi_{jv}(t)$$

where $\xi_{jv}(t)$ are centred random functions members

$$\sum_{i,j,v} a_{ij}\,W_{kv}(D)\,\Delta X_i\,\xi_{jv}(t) \tag{9}$$

play the same kind of negative role as members (8). In a quasi-stationary regime the influence of these members may be decreased to the same extent, as the influence of members (8), since the correlation times of function $\xi_{jv}(t)$ are compared with the correlation times of function $\delta X_{jw}\,\delta X_v$. In a quasi-stationary regime one neglects the influence of members (9).

And so, the general equations (5) give up their place to the following equations of a quasi-stationary regime of the system under consideration.

$$D\Delta X_k^*=\sum_{i,j,v} a_{ij}\,R_{jv}(\tau_v-\tau_j+\delta\tau_v)\,W_{kv}(D)\,\Delta X_i^*(t-\tau_i)+\delta f_k(t)-DX_{kl} \tag{10}$$

$$(k=1,2,\ldots,n)$$

These general equations of a quasi-stationary regime are simplified in concrete, particular cases.

First of all it is noted that the correlation times of search signals are small, due to the presence of high frequency filters. Therefore, for a typical case, when the delay times τ_j are not identical it may be assumed

$$R_{jv}(\tau_v-\tau_j+\delta\tau_v)=\begin{cases}0 & \text{for } j\neq v\\ R_v(\delta\tau_v) & \text{for } j=v\end{cases} \tag{11}$$

It should be noted that the same correlations take place also at strictly identical delay times $\tau_v=\tau_j$, but not correlated search constituents. In practice, non-correlated natural search constituents may be obtained by means of installing instead of high frequency filters, band filters with non-overlapping passing bands. The shortcoming of this method is the considerable lowering of the level or efficiency of the utilized search elements, especially in multi-instrument systems (n dimensional).

At condition (11) eqns (10) take the shape

$$D\Delta X_k^*+\sum_{i=1}^{n} W_{ki}^c(D)\,\Delta X_i^*(t-\tau_i)=\delta\phi_k-\dot{X}_{kl} \tag{12}$$

where

$$W_{ki}^c(b)=\sum_{v=1}^{n} a_{iv}R_v(\delta\tau_v)\,W_{kv}(D)$$

It is also possible to introduce transfer functions of a closed-loop system, then

$$\Delta X_k^*=\frac{1}{\Delta(D)}\sum_{v=1}^{n}(-1)^{k+v}\Delta_{kv}(D)(\delta\phi_v-\dot{X}_{vl}) \tag{13}$$

where

$$\Delta(D)=\begin{vmatrix} D+W_{11}^c(D)\,e^{-\tau_1 b}\ldots W_{1n}^c e^{-\tau_n b}\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ W_{n1}^c(b)\,e^{-\tau_1 b}\ldots\ldots D+W_{nn}(D)\,e^{-\tau_n b}\end{vmatrix}$$

$\Delta_{kv}(D)$ is the determinant, obtained from $\Delta(D)$ by crossing out 'K' column, 'v' line.

The roots of a characteristic equation

$$e^{-(\tau_1+\ldots+\tau_n)\lambda}\begin{vmatrix}\lambda e^{\tau_1\lambda}+W_{11}^c(\lambda)\ldots W_{1n}^c(\lambda)\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ W_{n1}^c\ \ldots\ldots\ldots\ldots\lambda e^{\tau_n\lambda}+W_{nn}^c(\lambda)\end{vmatrix}=0 \tag{14}$$

are simply determined in case when times τ_v are practically equal, the synchronous detector filters are neutral and possess identical transfer functions:

$$R_v(\delta\tau_v)\,W_{kv}(D)=\begin{cases}0 & \text{for } v\neq k\\ W(D) & \text{for } v=k\end{cases}$$

In this case the characteristic equation (14) breaks up into n equations of

$$\frac{\lambda e^{\tau\lambda}}{W(\lambda)}=-\frac{1}{C_v^2} \tag{15}$$

where C_v is the semi-axis of the determining ellipsoid

$$\sum_{i,k=1}^{n} a_{ik}\,\Delta X_i \Delta X_k=1$$

If by decreasing the gain $W(0)$ the roots of the characteristic equation are made so small, that $e^{\tau\lambda}\approx 1$, $W(\lambda)\approx W(0)$, then, in accordance with (15)

$$\lambda=-\frac{W(0)}{C_v^2}<0$$

and similar slow processes of extremum control always possess monotonous stability. However, at small gains the extremum control time or the self-adjusting time is great and the errors considerable.

Errors produced by drifting of the extremum with constant speed equal

$$\Delta X_k^*=-\frac{1}{\Delta(0)}\sum_{v=1}^{n}(-1)^{k+v}\Delta_{kv}(0)\,\dot{X}_{vl} \tag{16}$$

It is noted that the value

$$(-1)^{k+v}\frac{\Delta_{kv}(0)}{\Delta(0)}$$

equals the area bound by the curved weighting function $K_{kv}(t)$ of a closed-loop system, corresponding to the transfer function

$$\frac{\Delta_{kv}(D)}{\Delta(D)}$$

$$(-1)^{k+v}\frac{\Delta_{kv}(0)}{\Delta(0)}=\int_0^\infty K_{kv}(t)\,dt=T_{kv} \tag{17}$$

Values T_{kv} have time dimensions and will be called 'areas of weighting functions'. If the system of extremum control is in general disconnected at $\Delta X_k^*(0)=0$, *To* $W_{ki}{}^c=0$,

$$\delta f_k=0\,u$$

$$\Delta X_k^*=-\int_0^t \dot{X}_{kl}\,dt=-\dot{X}_{kl}\,t$$

where we consider $\dot{X}_{kl}=\text{const}$.

With a disconnected system of extremum control, deviation ΔX_k^* increases and in time I_k exceeds the permissible value ΔX_{kg}^* where

$$\Delta X_{kg}^*=-T_k\dot{X}_{kl} \tag{18}$$

Time intervals T_k are called adjustment-loss times, as distinct from the general adjustment-loss time, mentioned above.

From (16) and (17) it follows that

$$\Delta X_k^*=-\sum_{v=1}^{n}T_{kv}\,\dot{X}_{vl}$$

Errors ΔX_k^*, produced by constant drifting of the extreme point in a closed system, naturally, must not exceed ΔX_{kg}^*.

It follows from this that the weighting function areas must satisfy correlations

$$\left|\sum_{v=1}^{n}T_{kv}\,\dot{X}_{vl}\right|<T_k|X_{kl}|=|\Delta X_{kg}^*| \tag{19}$$

Assuming in particular

$$\dot{X}_{1l}=\ldots=\dot{X}_{k-1l}=\dot{X}_{k+1l}=\ldots=\dot{X}_{nl}=0$$

(moreover ΔX_{vg}^* remain final values) one obtains

$$|T_{kk}|<T_k$$

i.e. the weighting function areas must be smaller than the adjsutment-loss times.

The curtailment of the weighting function areas (decrease of static errors) may be attained by means of increasing of the amplification. However, the increase of gains, beginning with certain values, leads to loss of stability of the extreme loop.

The increase in critical values of the gains and curtailment of times of transient processes of the extremum loop requires the diminution of transient displacements τ_v between points of action of controlling elements and control points of output parameters in the industrial process itself.

The control of output parameters may be realized at the output of the whole industrial process [*Figure 4 (a)*], in the intermediate points [*Figure 4 (b)*], and at the output [*Figure 4 (c)*].

From the viewpoint of lag decreasing and possibility of time curtailment of transient processes, a circuit having parameter control in the intermediate points has a decisive advantage over a scheme with control of output [*Figure 4 (a)*] final since it corresponds to the arrangement of information transmitters in the immediate vicinity of the controlling elements. However, this circuit also has one essential drawback.

The quality index extremum, calculated on the basis of measurements in intermediate points may not correspond to the extremum of output quality at the industrial process output.

A more perfect circuit is the combined type [*Figure 4 (c)*] where the control in the intermediate points is combined with the output parameter control at the industrial process exit. In this circuit the signals of the parameter transmitters of the finished production pass through low frequency narrow-band filters Q, for instance integrating elements, after which they are added to the signals of corresponding transmitters, controlling the output parameters in the intermediate points. This circuit conserves the quick action of circuit *4 (b)* and at the same time possesses the accuracy of control of slow changing parameters, near to the accuracy of control of circuit *4 (a)*.

However, the above extremum control system even with an improved informational section has a limited general application.

In fact, as seen from eqn (12) the dynamics of quasi-stationary processes of extremum control in this system depends on coefficients a_{iv} of quadratic shape of the quality index, depends on time lag τ_v, errors of simulation of this delay $\delta\tau_v$ and intensity of search elements.

For some industrial processes these parameters may be considered permanent, for others they are subject to comparatively slow random variations.

For the first processes the extremum control loop once adjusted maintains its efficiency for a long time. In the second case the extremum control loop itself needs periodic adjustment, accomplished by changing the transfer functions $W_{kv}(D)$ of the filters or just their gains $W_{kv}(0)$, and also, perhaps, the time delays.

The necessity of adjustment arises due to the fact that, even though the stability of slow processes of the extremum control is maintained in a wide range of variations a_{iv}, $R_v(\delta\tau_v)$, a guarantee of the necessary quality of the extremum control is possible only by a suitable selection of transfer function filters.

To this must be added, that even in those processes where parameters a_{iv}, τ_v, $R_v(\delta\tau_v)$ remain unchanged the initial adjustment requires either *a priori* knowledge of these parameters, or their experimental determination.

Raising the chance of general acceptance of extremum control systems with continuous industrial processes, in the sense of volume decrease of necessary *a priori* information, may be attained at the expense of parametric extremum adjustment of the basic extremum loop.

Self-Adjusting System with Parametric Extreme Adjustment of the Basic Loop

To realize an extremum adjustment of the basic control, it is desirable to select the adjustment parameters of this loop and the quality index so that the latter shall be the only extremum in the working portion of possible variations of adjustment parameters.

As an adjustment index of the basic loop, it is natural to select the mean value, more accurately, the mathematical expectation of the same production quality index Q, which is utilized in the basic extremum loop.

Moreover, in accordance with (2)

$$M[Q]=Q_l+\frac{1}{2}\sum_{i,j=1}^{n} a_{ij}M[\Delta X_i(t-\tau_i)\Delta X_j(t-\tau_j)] + M[\delta Q_f]$$

it is possible to show, if the errors of simulation lag are so restricted, that

$$R_v(\delta\tau_v)>0 \quad (v=1,2,\ldots,n)$$

$$\begin{vmatrix} d_{11} \ldots d_{1n} \\ \cdots\cdots\cdots \\ \cdots\cdots\cdots \\ d_{n1} \ldots d_{nn} \end{vmatrix} > 0 \qquad (20)$$

where $d_{uk} = W_{uk}(0)$.

Then the estimation $M[Q]$ always has an extremum by the adjustment parameters d_{vk} whereupon this extremum is the only one in region (20).

Taking into account the availability of the single extremum by the adjustment parameters and the general principle of extremum control, it is easy to lay out a control system of an industrial process with seelf-adjustment of the basic loop.

This chart is shown in *Figure 5*. The basic loop of the extremum control of the industrial process is here similar to the one previously examined. The difference is only in the presence of multiplier 'matrix' links, which realize varying transfer numbers d_{vk}.

Besides the basic loop the system has a loop of extremum adjustment of the transfer numbers matrix.

It is assumed that the transfer numbers a_{iv} of the industrial process change slowly in time even as compared with quasi-stationary processes of the basic extremum loop. More accurately, it is assumed, that the time of substantial change of the transfer numbers a_{iv} is considerably greater than the general adjustment-loss time T (see above). It is noted that, in principle, forced self-adjusting processes of transfer numbers are also possible. However, the dynamics of forced processes is complex, and for their organization it is not enough only to have the existence of an extremum of appraisal $M[a]$ by the transfer numbers of the basic loop. Thus, as a typical regime of the system operation with two extremu loops, a regime with the following grading of process flow speeds is assumed: (*a*) Search elements in the basic loop (the frequency processes); (*b*) working processes in the basic loop; (*c*) search of oscillations in the loop of extremum self-adjustment of the transmitting numbers, and (*d*) working processes of the extremum self-adjustment of transfer numbers (the frequency processes).

At the above grading of process flow speeds, both the processes in the basic loop, as well as the processes in the self-adjusting loop of transfer numbers are quasi-stationary. The dynamics of working processes, moreover, are near to the dynamics of ideal gradient systems (4). From this position and presence of extremum of transfer numbers d_{vk} it follows, that upon fulfillment of weak conditions (20) a quasi-stationary process of self-adjustment of the basic loop is always stable.

The above control system has considerable universal acceptance. By joining it with a plant (industrial process) with little known characteristics, the system matches automatically transfer numbers corresponding to the quality index extremum in the framework of the given structure of the system.

A further increase in 'flexibility' or universality of the system is made by introducing extremum adjustment of the delay simulator, extremum adjustment of the filter time constants and others. However, all of this involves further complexity of the system.

The possibility of Extremum Control of Non-automated Control and Adjustment

The main difficulties in introduction of extremum control of industrial processes at the present stage are connected with the complete automation of output parameter control and complete automation of machine adjustment, securing the industrial process. The technology of most industrial processes even continuous, did not yet reach the level at which it is possible to achieve continuous automatic control and adjustment. Therefore, it is of great interest to find the means of extremum control for discrete semi-automatic or hand control and adjustment.

The general algorithm of the extremum control and the computing section of the control system may, moreover, remain the same, as in a continuous automatic system.

Estimation of the information output capacity of the measuring points, necessary for transmission of the search elements, indicates that for processes with considerable adjustment-loss times a non-automated control is possible. For such processes, a periodic hand adjustment of machines is also

possible, which guarantees the industrial process taking place. In this case the output signals of synchronous detectors are transmitted to the integrated indicators. Operators (adjusters), guided by the indicators of these devices, are periodically correcting the adjustment of the machines. With this type of organization of the extremum control the control system itself becomes a computer, either digital or analogue, equipped with input and output arrangements. The closing of the loop of an extremum control is here accomplished by men supervisors and operators.

References

[1] KRASOVSKI, A. A. *Some Conditions of Application of Self-adjusting Control Systems with Continuous Industrial Processes.* 1961. Moscow; Izvestia Academy of Sciences U.S.S.R.

[2] COWDEN, D. T. *Statistical Methods in Quality Control.* 1957.

[3] PERLMAN, I. I. Statistical automats, relay type and some methods of their investigation. *Automat. Telemech., Moscow* 22 (1961) 6

[4] KRASOVSKI, A. A. *Dynamics of Continuous Systems of Extreme Regulation, Based on Gradient Method.* 1959. Moscow; Izvestia Academy of Sciences U.S.S.R.

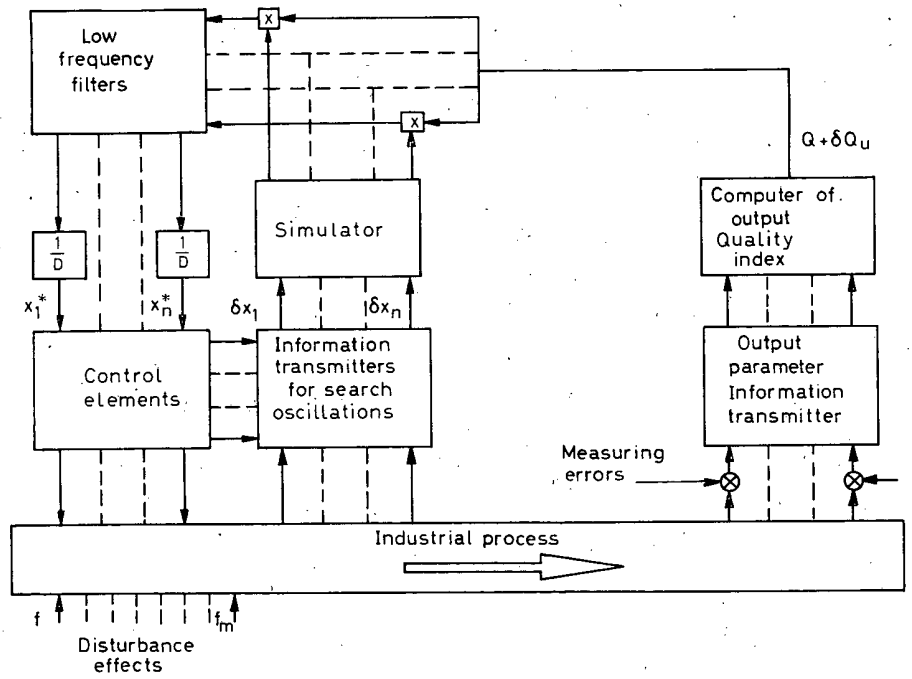


Figure 1. Layout of an extreme control system of an industrial process with a single criteria of production quality

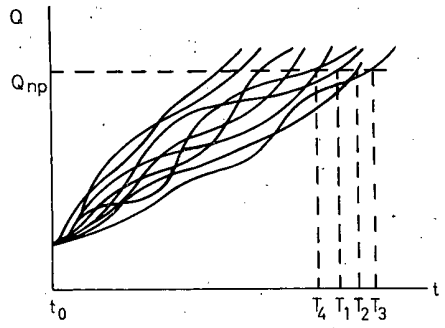


Figure 2. For determination of conception of general disadjustment time

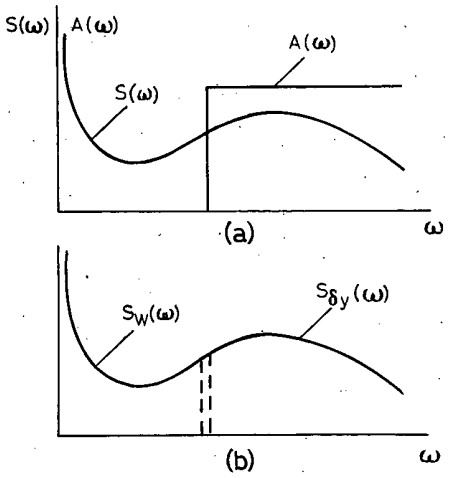


Figure 3. Separation of an independent search component by an ideal high frequency filter

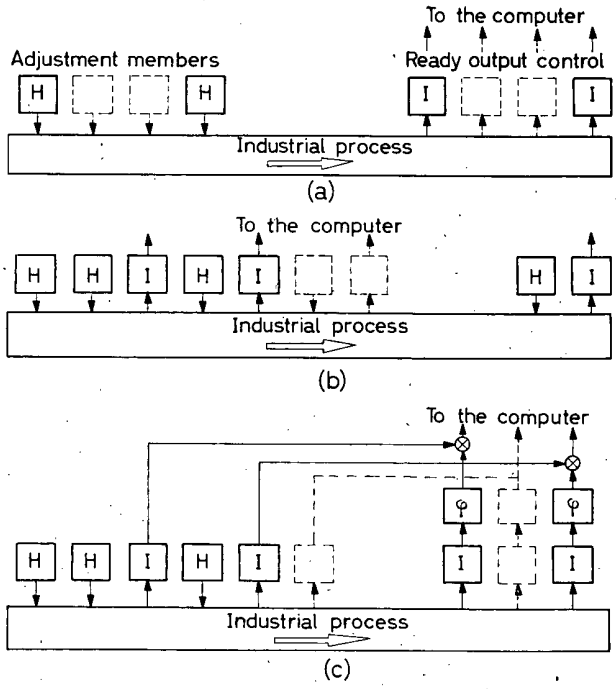


Figure 4

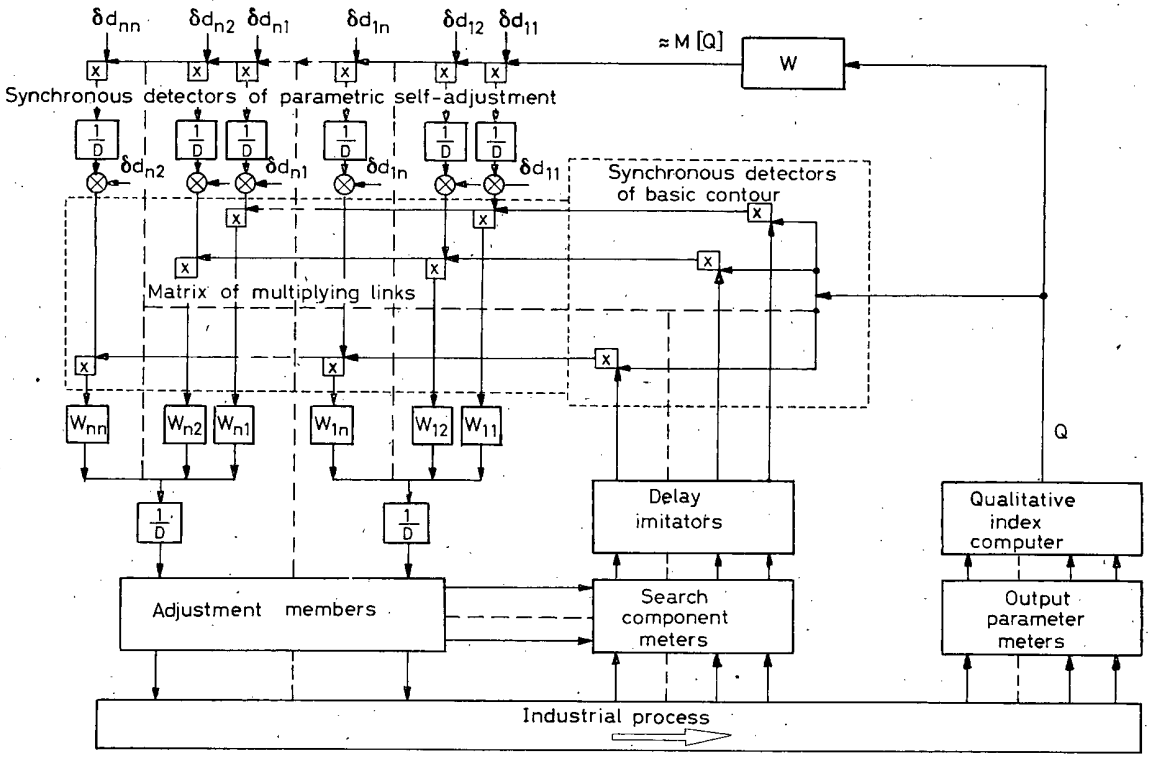


Figure 5

Invariance of Sampled-data and Adaptive Sampled-data Systems

V. M. KUNTSEVICH and Yu. V. KREMENTULO

One of the important scientific trends in the theory of automatic control is the theory of the construction of systems on the basis of compensation of the influence of disturbances, or the theory of invariance of the control led value.

As is known, however, the invariance theory was recently used extensively only for ordinary continuous control systems[1–7]. Attempts were made in a number of works[8–15] to extend the general principles of this theory to sampled-data control systems, but there has not yet been any full and systematized statement of the invariance theory for such systems. That said above relates in a still greater degree to adaptive systems in general and sampled-data systems of this type in particular. Since adaptive systems are a special type of non-linear systems, then, as will be shown below, the introduction of compounding disturbance links makes it possible not only to improve the quality of systems when compensating the influence of disturbances, but also to extend the stability region of these systems.

The authors consider the main aim of their paper is to demonstrate the fact that the sampled-data system analysis and synthesis methods expounded below can serve as the basis for the construction of control systems with considerably greater accuracy than existing systems.

Henceforward the following constraints and assumptions will be accepted: (*a*) synchronous sampled-data systems with amplitude modulation are considered; (*b*) the sampling period T is constant; (*c*) the pulse element is ideal; (*d*) the equations are written in deviations; and (*e*) initial conditions are zeroth.

Since sampled-data systems of fairly complex structure will be considered, the consideration will begin with the method of solving the equations of multi loop sampled-data systems.

Sampled-data Systems

Multiloop Sampled-data Systems Equations

A number of works[16–23] have been concerned with the compilation and solution of the equations of sampled-data systems. The solution of the equations of multiloop sampled-data systems is given in the most general and convenient form by Burshtein[17]. The method suggested below has features in common with Burshtein's method, but allows one to avoid a number of intermediate operations and to simplify the calculations.

In the most general form the equation for the kth coordinate of a multiloop sampled-data system can be written thus:

$$x_k(s) = \sum_{i=1}^{n} W_{ki}(s)\,x_i(s) + \sum_{i=1}^{n}\sum_{j=1}^{l_{ki}} b'_{kij}\,x_i^*(z)\,b_{kij}(s) + \sum_{i=1}^{m} R_{ki}(s)\,F_i(s) + \sum_{i=1}^{m}\sum_{j=1}^{P_{ki}} C'_{kij}\,F_i^*(z)\,C_{kij}(s) \qquad (1)$$

where x_k, x_i are the coordinates of the system, F_i the externa disturbances, n, m the number of selected coordinates and external disturbances respectively, l_{kij}, P_{kij} the number of parallel links (pulse-continuous) between the coordinate x_k and x_i and the coordinate x_k and the external effect F_i respectively; W, b, b', c, c' and R are the corresponding transfer functions, shown in *Figure 1*, which depicts part of a multiloop sampled-data system (kth node).

If one takes into consideration the additional coordinates:

$$\begin{aligned} &b'_{111}(s)\,x_1(s) = x_{n+1}(s)\\ &\cdots\cdots\cdots\\ &b'_{11l_{11}}(s)\,x_1(s) = x_{n+l_{11}}(s)\\ &b'_{1n1}(s)\,x_n(s) = x_{n+l_{11}+\cdots l_{1(n-1)}+1}(s)\\ &\cdots\cdots\cdots\\ &b'_{1nl_{1n}}(s)\,x_n(s) = x_{n+l_{11}+\cdots+l_{1n}}(s), \text{ etc.} \end{aligned} \qquad (2)$$

then the equations of the multiloop sampled-data system can be given in an ordered form:

$$\begin{aligned} &\sum_{j=1}^{N} a_{1j}(s)\,x_j(s) + \sum_{j=1}^{N} a'_{1j}(s)\,x_j^*(z) = A_1(s)\\ &\cdots\cdots\cdots\\ &\sum_{j=1}^{N} a_{Nj}(s)\,x_j(s) + \sum_{j=1}^{N} a'_{Nj}(s)\,x_j^*(z) = A_N(s) \end{aligned} \qquad (3)$$

where

$$N = n + \sum_{k=1}^{n}\sum_{i=1}^{n} l_{ki}$$

is the full amount of coordinates of the system (including the additional ones),

$$A_k(s) = -\left[\sum_{i=1}^{m} R_{ki}(s)\,F_i(s) + \sum_{i=1}^{m}\sum_{j=1}^{P_{ki}} C'_{kij}\,F_i^*(z)\,C_{kij}(s)\right],$$

$$a_{kj}(s) = W_{ki}(s); \quad a_{kk}(s) = W_{kk}(s) - 1; \quad a'_{Kj}(s) = b'_{Kij}(s)$$

and are numbered in accordance with (2).

System (3) formally contains N equations with $2N$ unknowns. $x_j(s)$ and $x_j^*(z)$. As in ref. 17, the terms containing transforms of the coordinates will be transferred to the right-hand side. The resultant system will be solved relative to the arbitrary coordinate $x_j(s)$. This gives:

$$x_j(s) = \frac{\Delta x_j(s)}{\Delta(s)} \qquad (4)$$

where

$$\Delta(s) = \begin{vmatrix} a_{11}(s), \ldots, a_{1N}(s)\\ \cdots\cdots\cdots\\ a_{N1}(s), \ldots, a_{NN}(s) \end{vmatrix} \qquad (5)$$

is the common determinant of a purely continuous system.

Eqn (6) *

Determinant (6) may be presented in the form:

Eqn (7) **

The first of the determinants entering into (7) will be denoted by $\Delta_A^j(s)$, and the remainder by $\Delta^j_{x_k^*}(s)$. Bearing in mind the notation adopted:

$$x_j(s)=\frac{\Delta_A^j(s)}{\Delta(s)}-\sum_{k=1}^{N} x_k^*(z)\left(\frac{\Delta^j_{x^*_k}(s)}{\Delta(s)}\right) \tag{8}$$

Subjecting (8) to a z transform and cancelling out like terms, the following relation is obtained:

$$x_j^*(z)\left[1+\left(\frac{\Delta^j_{xj^*}}{\Delta}\right)^*(z)\right]=\left(\frac{\Delta_A^j}{\Delta}\right)^*(z)-\sum_{k=1}^{N} x_k^*(z)\left(\frac{\Delta^j_{x^*_k}}{\Delta}\right)^*(z);\quad (k\neq j) \tag{9}$$

Thus the initial system (3) can immediately be raised to a full system by equations of type (9). The full system of equations of a multiloop sampled-data system has the form:

$$\begin{aligned}
&\sum_{j=1}^{N} a_{1j}(s)\,x_j(s)+\sum_{j=1}^{N} a'_{1j}(s)\,x_j^*(z)=A_1(s)\\
&\cdots\cdots\cdots\cdots\cdots\\
&\sum_{j=1}^{N} a_{Nj}(s)\,x_j(s)+\sum_{j=1}^{N} a'_{Nj}(s)\,x_j^*(z)=A_N(s)\\
&\left[1+\left(\frac{\Delta^1_{x^*_1}}{\Delta}\right)^*(z)\right]x_1^*(z)\\
&+\sum_{j=2}^{N}\left(\frac{\Delta^1_{x^*_j}}{\Delta}\right)^*(z)\,x_j^*(z)=\left(\frac{\Delta_A^1}{\Delta}\right)^*(z);\quad (j\neq N)\\
&\cdots\cdots\cdots\cdots\cdots\\
&\sum_{j=1}^{N-1}\left(\frac{\Delta^N_{x^*_j}}{\Delta}\right)^*(z)\,x_j^*(z)+\left[1+\left(\frac{\Delta^N_{x^*_N}}{\Delta}\right)(z)\right]x_N^*(z)=\left(\frac{\Delta_A^N}{\Delta}\right)^*(z);
\end{aligned} \tag{10}$$

When writing the determinants forming part of (10), the following symbolization is accepted. The upper index shows which column of the common determinant Δ is subject to substitution, while the lower index indicates substitution by coefficients for particular variables. Thus, $\Delta^k x^*_j$ means that the kth column of the common determinant Δ is to be replaced by coefficients at the jth discrete coordinate.

System (10) can be solved, relative to the coordinates of interest, by ordinary, algebraic methods. Sampled-data systems with various types of link will now be considered.

Sampled-data Systems with Continuous Compounding Links

An automatic control system with one pulse element, which can be described by a system of three linear equations with constant coefficients, is studied. The block diagram of the system is given in *Figure 2*, which also shows the transfer functions of both the main loop and the additional links.

The initial system of equations is:

$$\begin{aligned}
&\varphi(s)+0-W_{\varphi\mu}W_{\varphi\lambda}(s)+0=(s)\,\lambda(s)\\
&-W_{v\varepsilon}(s)\,W_{\varphi_1\varphi}(s)\,\varphi(s)+v(s)-W_{v\mu}(s)\,\mu(s)+0\\
&=W_{v\lambda}(s)\,\lambda(s)+W_{v\varepsilon}(s)\,\psi(s)\\
&-W_{\mu\varphi}(s)\,\varphi(s)+0+\mu(s)-W_{\mu v}(s)\,v^*(z)\\
&=W_{\mu\lambda}(s)\,\lambda(s)+W_{\mu\psi}(s)\,\psi(s)
\end{aligned} \tag{11}$$

In accordance with the method expounded above, this system is made into a full one by the deficient equation:

$$\left[1+\left(\frac{\Delta^2_{v^*}}{\Delta}\right)(z)\right]v^*(z)=\left(\frac{\Delta_A^2}{\Delta}\right)^*(z) \tag{12}$$

Henceforward, only programme and servosystems will be considered; hence, in (11), $\lambda(s)=0$.

From (11) and (12) one can easily find an expression for the controlled coordinate in which one is interested.

$$\varphi(s)=K_2(s)\,\psi(s)+\frac{K_3(s)}{1-K_7^*(z)-K_3K_6^*(z)}\{[(K_5+K_2K_6)\,\psi]^*(z)\} \tag{13}$$

where

---

* $$\Delta_{x_j}(s)=\begin{vmatrix} a_{11}(s),\ldots,a_{1j-1}(s);\ A_1(s)-\sum_{j=1}^{N} a'_{1j}(s)\,x_j^*(z);\ a_{1j+1}(s),\ldots,a_{1N}(s)\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ a_{N1}(s),\ldots,a_{Nj-1}(s);\ A_N(s)-\sum_{j=1}^{N} a'_{Nj}(s)\,x_j^*(z);\ a_{Nj+1}(s),\ldots,a_{NN}(s)\end{vmatrix} \tag{6}$$

** $$\begin{aligned}\Delta_{x_j}(s)=&\begin{vmatrix} a_{11}(s),\ldots,a_{1j+1}(s);\ A_1(s);\ a_{1j+1}(s),\ldots,a_{1N}(s)\\ a_{N1}(s),\ldots,a_{Nj+1}(s);\ A_N(s);\ a_{Nj+1}(s),\ldots,a_{NN}(s)\end{vmatrix}\\ &-\sum_{k=1}^{N} x_k^*(z)\begin{vmatrix} a_{11}(s),\ldots,a_{1j-1}(s);\ a'_{1k}(s);\ a_{1j+1}(s),\ldots,a_{NN}(s)\\ a_{N1}(s),\ldots,a_{Nj-1}(s);\ a'_{Nk}(s);\ a_{Nj+1}(s),\ldots,a_{NN}(s)\end{vmatrix}\end{aligned} \tag{7}$$

$$K_2(s)=\frac{W_{\varphi\mu}(s)\,W_{\mu\psi}(s)}{1-W_{\varphi\mu}(s)\,W_{\mu\varphi}(s)};\quad K_3(s)=\frac{W_{\mu v}(s)\,W_{\varphi\mu}(s)}{1-W_{\varphi\mu}(s)\,W_{\mu\varphi}(s)};$$

$$K_5(s)=W_{v\varepsilon}(s)+W_{v\mu}(s)\,W_{\mu\psi}(s);$$

$$K_6(s)=W_{\varphi_1\varphi}(s)\,W_{v\varepsilon}(s)+W_{v\mu}(s)\,W_{\mu\varphi}(s);$$

$$K_7(s)=W_{v\mu}(s)\,W_{\mu v}(s)$$

Conditions of Absolute Invariance. The condition of absolute invariance for servo and programme systems is:

$$\varphi(s)=K_2(s)\,\psi(s)+\frac{K_3(s)}{1-K_7^*(z)-K_3K_6^*(z)}\{[(K_5+K_2K_6)\,\psi]^*(z)\}=\psi(s)$$

or

$$-\varepsilon(s)=K_2'(s)\,\psi(s)+\frac{K_3(s)}{1-K_7^*(z)-K_3K_6^*(z)}\{[(K_5+K_2K_6)\,\psi]^*(z)\}=0 \tag{14b}$$

where $\varepsilon(s)=\psi(s)-\varphi(s)$ is the system error of the system; $K'_2(s)=K_2(s)-1$.

The basic differences between the conditions of invariance for continuous and sampled-data systems is emphasized. While in continuous systems the conditions of absolute invariance do not depend on the form of ψ, and are determined only by the parameters of the components of the system, in the sampled-data system under consideration, these conditions (14) essentially depend on the form of the input signal ψ.

It can be shown that the condition of absolute invariance physically signifies the equality to zero of the sum of the individual components of the coordinate ε produced both as a result of the direct effect ψ upon the system, and also on account of the effect *via* the additional (compounding) links.

Invariance Conditions for Discrete Moments of Time. The invariance conditions (14) were obtained from the requirement of the equality to zero of coordinate ε at any moments of time. One may pose a less rigid requirement—the equality to zero of ε at the sampling instants, i.e.,

$$\varepsilon[nT]=0 \tag{15}$$

The conditions under which (15) is satisfied are called 'conditions of invariance for discrete moments of time'. If (14) is subjected to a z transform, then the problem is solved at first sight. However, it is easy to show that the invariance conditions for discrete moments of time as well, will depend upon ψ.

An attempt is made to obtain the conditions, independent of ψ. Both parts of (14b) are multiplied by

$$\frac{K_5(s)+K_2(s)\,K_6(s)}{K_2'(s)}$$

and then subjected to a z transform.

$$-\left(\frac{l}{K_2'}\,\varepsilon\right)^*(z)=(l\psi)^*(z)+\left(\frac{K_3 l}{K_2'}\right)^*(z)\frac{(l\psi)^*(z)}{1-K_7^*(z)-K_3K_6^*(z)} \tag{16}$$

is obtained, where $l(s)=K_5(s)+K_2(s)\,K_6(s)$.

By equating the right-hand side of (16) to zero, the following invariance conditions are obtained for discrete moments of time:

$$1-K_7^*(z)+\left[\frac{K_3(K_5+K_6)}{K_2'}\right]^*(z)=0 \tag{17}$$

The conditions of absolute invariance for a similar continuous system (i.e., a system having the same structure) can be given in the form:

$$1-K_7(s)+\frac{K_3(s)[K_5(s)+K_6(s)]}{K_2'(s)}=0 \tag{18}$$

If (18) is subjected to a z transform, eqn (17) is obtained, i.e., *the introduction of a pulse element into an absolutely invariant continuous system does not impair the conditions of invariance for discrete moments of time for the so-called 'fictitious coordinate'*

$$\varepsilon_\varphi(s)=\frac{l(s)}{K_2'(s)}\,\varepsilon(s)$$

As shown by Krementulo[10] from the equality to zero of $\varepsilon_\varphi[nT]$, there still does not follow the equality to zero of $\varepsilon[nT]$. The additional conditions will be given, under which $\varepsilon[nT]=0$, and does not depend on the form of ψ. (14b) is subjected to a z transform, and then $1-K_7^*(z)$, found from (17), is substituted:

$$-\varepsilon^*(z)=K_2'\psi^*(z)-\frac{K_3^*(z)}{\left\{\frac{K_3(K_5+K_6)}{K_2'}\right\}^*(z)+K_3K_6^*(z)}\times\{K_5\psi^*(z)+[(K_2'+1)\,K_6\psi]^*(z)\} \tag{19}$$

The additional condition:

$$\left[\left(\frac{K_5+K_6}{K_2'}+K_6\right)K_2'\psi\right]^*(z)=\left(\frac{K_5+K_6}{K_2'}+K_6\right)^*(z)\,K_2'\psi^*(z) \tag{20}$$

Condition (20) is satisfied if $[(K_5+K_6)/K_2']+K_6$ contains proportional components or components with a pure time lag. From (20) and (17) can be found the transfer functions of continuous compounding links.

Sampled-data Systems with Discrete Compounding Links

A brief examination will be made of the properties of a typical sampled-data servo-system, the block diagram of which is given in *Figure 3*. The expression of the system error ε is:

$$\varepsilon^*(z)=\frac{1-W_{\mu\psi}^*(z)\,W_{\varphi\mu}^*(z)}{1+W_{v\varepsilon}^*(z)\,W_{\varphi\mu}^*(z)}\,\psi^*(z) \tag{21}$$

The condition of invariance at discrete moments of time is:

$$W_{\mu\psi}^*(z)=\frac{1}{W_{\varphi\mu}^*(z)} \tag{22}$$

In the general case, $W_{v\varepsilon}^*(z)$ and $W_{\varphi\mu}^*(z)$ are the ratio of polynomials according to the positive powers of z, the power of the numerator being less than that of the denominator. Since $W_{\mu\varphi}^*(z)$ must be inverse to $W_{\varphi\mu}^*(z)$, then it cannot

be physically realized (advancing components are required for this).

It is important to note that the introduction of the link $W^*_{\mu\psi}(z)$ and the satisfaction of the invariance condition (22) do not alter the characteristic equation of the system:

$$K^*_0(z)P^*(z)+K^*_1(z)Q^*(z)=0;$$
$$\left(\frac{K^0_*(z)}{K^*_1(z)}=W^*_{v\varepsilon}(z);\ \frac{P^*(z)}{Q^*(z)}=W^*_{\varphi\mu}(z)\right) \qquad (23)$$

and therefore do not influence the stability of the system.

Examples were given by Kuntsevich[12] to show that even in those cases when $W^*_{\mu\psi}(z)$, obtained from condition (22) cannot be realized, provided it is selected in a particular way, it is possible to increase considerably the accuracy of a sampling servosystem.

When for any reasons it is inconvenient or impossible to introduce the compounding link $W^*_{\mu\psi}(z)$, one may introduce into the system additional links, equivalent to the direct compounding link $W^*_{\mu\varphi}(z)$. Eqn (21) can be brought to the form:

$$\varepsilon^*(z)=\frac{1}{1+W^*_{v\varepsilon}(z)\,W^*_{\varphi\mu}(z)}\,\psi^*(z)$$
$$-\frac{W^*_{\mu\psi}\,W^*_{\varphi\mu}(z)}{1+W^*_{v\varepsilon}(z)\,W^*_{\varphi\mu}(z)}[\varepsilon^*(z)+\varphi^*(z)] \qquad (24)$$

It is not difficult to see that (24) is met by the scheme shown in *Figure 3 (b)*.

If (22) is satisfied, then the condition of absolute invariance has the form:

$$\frac{\psi(s)}{\psi^*(z)}=\frac{W_{\varphi\mu}(s)}{W^*_{\varphi\mu}(z)} \qquad (25)$$

The latter equality can be satisfied only in some particular cases, and, as shown by Krementulo[11], requires the inclusion of advancing components if $\psi[0]=0$.

Sampled-data Systems With Pulse-continuous Compounding Links

In this section a servosystem will be used as an example to show that when pulse-continuous links are used it is in principle possible to achieve absolute invariance in a combined sampled-data system.

Assume that the block diagram is predetermined, i. e., $W_{v\varepsilon}(s)$, $W_{\varphi\mu}(s)$ and $W_{\varepsilon\varphi}(s)$ are known. A compounding link with respect to the input signal ψ $W_{\mu\psi}(s)$ is introduced to improve the dynamic properties. The transfer function of this link has to be determined.

The expression for the system erroris:

$$-\varepsilon(s)=[W_{\mu\psi}(s)\,W_{\varphi\mu}(s)-1]\,\psi(s)$$
$$+\frac{W_{v\varepsilon}(s)\,W_{\varphi\mu}(s)}{1+W_{v\varepsilon}\,W_{\varphi\mu}\,W^*_{\varepsilon\varphi}(z)}[\psi^*(z)+W_{\mu\psi}\,W_{\varphi\mu}\,W_{\varepsilon\varphi}\,\psi^*(z)] \qquad (26)$$

Having equated $\varepsilon(s)$ to zero the condition of invariance of the system is obtained from which the transfer function of the compounding link can be determined:

$$W_{\mu\psi}(s)=\frac{1}{W_{\varphi\mu}(s)}+\frac{W_{v\varepsilon}(s)}{\psi(s)}[W_{\varepsilon\varphi}\,\psi^*(z)-\psi^*(z)] \qquad (27)$$

The signal of the compounding link $v_1(s)$ equals:

$$v_1(s)=\frac{\psi(s)}{W_{\varphi\mu}(s)}+W_{v\varepsilon}(s)[W_{\varepsilon\varphi}\,\psi^*(z)-\psi^*(z)] \qquad (28)$$

This signal can be realized with the aid of the scheme shown in *Figure 4 (b)*. In a similar continuous system, the compounding link with respect to ψ, chosen from the conditions of absolute invariance, equals:

$$W_{\mu\psi}(s)=\frac{1}{W_{\varphi\mu}(s)}+W_{v\varepsilon}(s)[W_{\varepsilon\varphi}(s)-1] \qquad (29)$$

It can be seen that for both the sampled-data and the continuous system the compounding link has one and the same structure and consists of identical components. The difference lies in the fact that in an absolutely invariant sampled-data system some of the components are connected up *via* additional pulse elements operating synchronously and in phase with the main one. What has already been said also holds in the case when real pulse elements are used.

Extremal Sampled-data Systems

Systems without Compounding Links

Today a large number of extremal sampled-data systems of various types are known, which have been studied by many scientists. But certain specific features of these systems remain unexplained. Of the known extremal sampled-data systems an analysis will be made on the basis of full and precise equations of dynamics of only one system which, as was shown in (29), provides the best tracking quality with continuous drift of the extremum, and whose properties are at the same time closest to those of a hypothetical system measuring the position of the extremum point without any errors.

As in most works, the controlled plant with extremal characteristics will be considered to be one which consists of a linear inertial component and an inertia-less component with extremal characteristics.

The equation of the non-linear component, taking into account the action of two kinds of disturbances (or two components of one and the same disturbance), which displace the extremum point, will be written in the form:

$$\varphi=-\alpha_3(x+\psi)^2+\lambda \qquad (30)$$

where φ is the index of the extremum, and ψ, λ are disturbances of an arbitrary kind, inaccessible for direct measurement by virtue of the conditions of the problem. Let the remaining equations of the extremal system (see *Figure 5*) in the absence of the components shown in *Figure 5* by the dotted line, be:*

$$x(s)=W_{xM}(s)\,M(s) \qquad (31)$$

where

$$M=\mu+m \qquad (31a)$$

* Since the system under review is non-linear, then strictly speaking, neither the ordinary nor the discrete Laplace transform is applicable to it. Therefore the final results will be obtained with the aid of a set of non-linear difference equations. To simplify things, the Laplace transform will only be used in application to the linear components.

$$m(s) = W_0(s)\, m_i^*(z) \tag{32}$$

where

$$m_i^*(z) = a'_M \frac{z}{1+z} \tag{32a}$$

$$\mu(s) = W_{\mu u}(s)\, u^*(z) \tag{33}$$

$$y_n = \Delta\varphi_{n-1}(-1)^n \tag{34}$$

$$u^*(z) = W_{uy}^*(z)\, y_{\Delta}^*(z) \tag{35}$$

Here (31) is the equation of the linear part of the plant, (32) the equation of the modulation circuit, (34) the equation of a controller with synchronous detector, (35) the equation of the correcting elements, (33) the equation of the servomotor and x, μ, φ, u and y the controlled coordinates.

Henceforward it is taken that the dynamic properties of the plant and the slope α_3 of the extremal characteristic are constant or quasi-constant.

The error of the system is denoted as:

$$e = \mu' + \lambda \tag{36}$$

and also the notations are introduced

$$x^*(z) = \mu'^*(z) + m'^*(z) \tag{37}$$

where

$$\mu'^*(z) = W_{xM}\, W_{\mu u}^*(z)\, u^*(z) \tag{37a}$$

$$m'^*(z) = W_0\, W_{\mu u}^*(z)\, m_i^*(z) \tag{37b}$$

On the basis of (37b) and (32, 32a), the modulating effect m'_n, scaled to the input of the non-linear element, can be represented in the form

$$m'_n = a_M \cos \pi n = a_M(-1)^n \tag{38}$$

where a_M is determined from the particular solution of the difference equation

$$a_M(-1)^n = a'_M\, W_{xM}\, W_0(E)(-1)^n \tag{39}$$

which is obtained following the replacement of (32) by the difference equation corresponding to it.

Solving jointly (30), (36), (37) and (38) gives

$$y_n = -2\, a_M \alpha_3 (e_n + e_{n-1}) + \Delta\lambda_{n-1}(-1)^n - \alpha_3 (e_n^2 - e_{n-1}^2)(-1)^n \tag{40}$$

From (40) it can be seen that the signal on the output of the component (34), apart from the useful component proportional to the error contains further additional terms, one of which $\Delta\lambda_{n-1}(-1)^n$ reflects the influence of the disturbance λ_n, and the third term shows that the measurement of the position of the system relative to the extremum point is not ideal.

Further replacing (35) and (33) by their corresponding difference equations, and solving then jointly with (40) and (37a), the equation of the dynamics of the system is obtained in the form of a non-linear difference equation with time-varying coefficients

$$[2\, a_M \alpha_3 W(E)(E+1) + E]\, e_n - \alpha_3 W(E)[e_{n+1}^2 - e_n^2)\cos \pi n] = \psi_{n+1} - W(E)[\Delta\lambda_n \cos \pi n]$$

$$\text{where } W(E) = W_{xM}\, W_{\mu u}(E)\, W_{uy}(E) \tag{41}$$

As was shown by Kuntsevich[29, 30], the non-linear eqn (41) has the peculiarity that at a particular correlation between the system parameters and the speed of variation of disturbances φ_n, λ_n the stability of the system is impaired, whereas analysis of the linearized equation obtained from (41), disregarding the non-linear terms (as done by Chang[25], Van-Neis[26] and Ivakhnenko[27]) does not permit one to detect this phenomenon. Therefore the feasibility of constructing an adaptive system, the error of which would be invariant in relation to ψ_n, λ_n, acquires particular interest, since it involves not only the improvement of the quality of the system, but also the increasing of its stability margin.

Invariance of Extremal Control Systems with Indirect Compounding Links

Since, by virtue of the conditions of the problem, the possibility of direct measurement of the signals ψ and λ is excluded, the possibility will be considered of using indirect compounding links with respect to ψ and λ similar to those considered above.

Consideration will first be given to the possibility of attaining invariance of system error at discrete moments of time, relative to ψ_n*.

From (41), (36), (42), and (42a) and also from *Figure 5*, it follows that

$$\psi^*(z) = e^*(z) - \mu'^*(z) \tag{42}$$

or

$$\psi^*(z) = e^*(z) - W_{xM}\, \mu^*(z) \tag{42a}$$

For the construction of the correcting link with respect to ψ in accordance with (42a), the variable μ'_n can be obtained with the aid of a model of the linear part of the controlled plant (see *Figure 5**). A signal proportional to e_n (or, more strictly, containing e_n) can be obtained on the output of an additional synchronous detector (see the part of *Figure 5* outlined by broken line), the equation of which is:

$$y'_n = \varphi_n(1)^n \tag{43}$$

Solving (30), (36) and (43) jointly, gives

$$y'_n = -2\, a_M \alpha_3 e_n - \alpha_3 (e_n^2 + a_M^2)(-1)^n + \lambda_n(-1)^n \tag{44}$$

For filtration of the parasitic quasi-periodic terms of signal (44) on the output of the detector in the network in *Figure 5*, a low-frequency filter is provided.

Taking this into account, the signal on the output of the additional control loop is written in the form

$$W_n \approx D(E)\, \psi_n \tag{45}$$

where

$$D(E) = 2\, a_M \alpha_3 W'_{\varphi}(E)\, W_K(E)$$

Omitting the intermediate operations, the equation of the dynamics of the system in *Figure 5*, with an additional control loop, is obtained, on the basis of the equations cited above and also eqn (45), in the form

$$[2\, a_M \alpha_3 W(E)(E+1) + E]\, e_n - \alpha_3 W(E)[(e_{n+1}^2 - e_n^2)\cos \pi n] = [1 - 2\, a_M \alpha_3 W_{xM} W_{\mu u}(E)\, W'_{\varphi}(E)\, W_K(E)]\, \psi_{n+1} - W(E)\, \Delta\lambda_n \cos \pi n \tag{46}$$

By equating to zero the operator comultiplier for ψ in the right-hand side of (46), an expression is obtained of the impulse

* It is noted that in contrast to ordinary servosystems, in which the input signal may also contain a noise which has to be suppressed as effectively as possible, the task of an extremal system in all cases is complete performance of signal ψ.

transfer function $W_K(E)$, which ensures the invariance of the system from ψ_n at discrete moments of time

$$W_K(E)=\frac{1}{2a_M\alpha_3}\frac{1}{W'_\varphi(E)W(E)} \tag{47}$$

From (46) it can be seen that the satisfaction of the conditions of invariance (47), and the presence of the filter in the compounding-link network (as distinct from the filter in the main network of the controller), do not alter the form and coefficients of the left-hand side of the equation of the dynamics of the system, i. e., do not directly influence the stability of the system.

When the required transfer function $W_K^*(z)$ is physically unrealizable, then, as for ordinary servosystems, a considerable improvement of accuracy (increasing of the degree of astatism) can be achieved by appropriate selection of the transfer function $W_K^*(z)$. An example is given in the Appendix of the method of selection of the coefficients of the transfer function $W_K^*(z)$.

In deriving the conditions of invariance (47), the quasi-periodic non-linear terms in (44) were disregarded in order to simplify the investigation. As follows from the example in the Appendix (see also *Figure 6*), the influence of these terms is in fact small.*

A brief examination will now be made of the possibility of minimization (or complete elimination) of the system error due to λ. From the equation of the system dynamics (46) and (40), it follows that for the predetermined structure the possibility of constructing a correcting link with respect to $\lambda(t)$ in a similar way as with respect to ψ, without constructing an analogue of the non-linear component, is excluded. By virtue of this, with the scheme structure adopted, only methods of minimizing the influenceof $\lambda(t)$ can be considered. One such method, based on the selection of the corresponding function $W^*_{uy}(z)$ was considered by Chang[25], Van-Neis[26] and Ivakhnenko[27]. The results obtained by Tou[24] may also be used here.

Appendix

Example—In *Figure 5* let

$$W_{xM}F(s)=\frac{\alpha_1}{\tau_1 s+1};\qquad W_{\mu u}(s)=\frac{\alpha_2}{s}$$

to which there corresponds

$$W_{xM}W^*_{\mu u}(z)=\frac{\alpha_1\alpha_2(1-d_1)z}{(z-1)(z-d_1)}$$

and further, let

$$W^*_{uy}(z)=\frac{B_1^*(z)}{B_2^*(z)}$$

where $B_1^*(z)$ and $B_2^*(z)$ are polynomials from z, $d_1=e^{-T/\tau_1}$.

It will be taken that

$$W'_\varphi(s)=\frac{1-e^{-sT}}{s}\frac{1}{\tau_2 s+1}$$

to which there corresponds

$$W'_\varphi(z)=\frac{1-d_2}{z-d_2};\qquad (d_2=e^{-T/\tau_2})$$

* The system in *Figure 5* was checked experimentally on an electronic analogue by A. A. Tunik, and the check confirmed the effectiveness of the introduction of indirect correction[31].

It is not difficult to see that in the given case the impulse transfer function $W_K^*(z)$, as determined from (47), which is required for attainment of the conditions of invariance, is physically unrealizable, and only the approximate satisfaction of the conditions of invariance can be spoken of; by virtue of this, $W_K^*(z)$ will be sought in the form of the series

$$W_K^*(z)=\sum_{i=1}^{K}C_i\left(\frac{z-1}{z}\right)^i \tag{48}$$

Denoting the left-hand side of equation (46) by $L(E)e_n$ in order to abbreviate the notation, one ean write it for $\Delta\lambda_n=0$ for the given example, bearing in mind (48), in the form:

$$\begin{aligned}L(E)e_n&=EB_2(E)\{-2a_M\alpha_1\alpha_2\alpha_3(1-d_1)(1-d_2)\\&\times[C_1\Delta\psi_n+C_2\Delta^2\psi_{n-1}+\ldots+C_K\Delta^K\psi_{n-K+1}]\\&+\Delta^3\psi_n+\Delta^2\psi_n[(1-d_1)+(1-d_2)]+\Delta\psi_n(1-d_1)(1-d_2)\}\end{aligned} \tag{49}$$

Provided

$$C_1=\frac{1}{2a_M\alpha_1\alpha_2\alpha_3} \tag{50}$$

the error from the first difference ψ_n is eliminated, since, when this is satisfied, the equation of the system adopts the form

$$\begin{aligned}&L(E)e_n\\&=EB_2(E)\{-2a_M\alpha_1\alpha_2\alpha_3(1-d_1)(1-d_2)[C_2\Delta^2\psi_{n-1}+\ldots\\&+C_K\Delta^K\psi_{n-K+1}]+\Delta^3\psi_n+(2-d_1-d_2)\Delta^2\psi_n\end{aligned} \tag{51}$$

Further taking

$$C_2=\frac{(2-d_1-d_2)}{2a_M\alpha_1\alpha_2\alpha_3(1-d_1)(1-d_2)} \tag{52}$$

and bearing in mind that

$$\Delta^i\psi_n-\Delta^i\psi_{n-1}=\Delta^{i+1}\psi_{n-1}$$

(51) can be rewritten in the form

$$\begin{aligned}&L(E)e_n\\&=EB_2(E)\{-2a_M\alpha_1\alpha_2\alpha_3(1-d_1)(1-d_2)[C_3\Delta^3\psi_{n-2}+\ldots\\&+C_K\Delta^K\psi_{n-K+1}]+\Delta^3\psi_n-C_2\Delta^3\psi_{n-1}\end{aligned} \tag{53}$$

from which it will be seen that, irrespective of the coefficients $W^*_{uy}(z)$ the error is eliminated from the second difference ψ_n. Since further increasing of the degree of astatism on account of the correcting link is impossible in the given example, $C_i=0$ will be taken for $i\geq 3$.

For quantitive evaluation of the quasi-periodic terms in (46), which have not been taken into account, in *Figure 6* the transient in an extremal system is plotted, taking into account these terms for $\psi_n=\beta n$, $\Delta\lambda_n=0$ for eqn (46).

For the transfer function of the components cited in the example under consideration and for $W^*_{uy}z=1$, the precise equation of the dynamics of the system has the form:

$$\begin{aligned}&A_0e_{n+3}+A_1e_{n+2}+A_2e_{n+1}+A_3e_n\\&=\alpha_\Sigma(1-d_1)[e^2_{n+2}-e^2_{n+1}+d_2(e^2_{n+1}-e^2_n)](-1)^n\\&+\alpha_\Sigma(1-d_1)(1-d_2)[e^2_{n+1}+e^2_n+2a^2_M](-1)^n\end{aligned} \tag{54}$$

where

$$A_0=1;\quad A_1=2a_M\alpha_\Sigma(1-d_1)-(1+d_1+d_2);$$

$$A_2 = 2a_M\alpha_\Sigma(1-d_1)(1-d_2)+d_1+d_2+d_1d_2;$$
$$A_3 = -d_1d_2-2a_M\alpha_\Sigma(1-d_1)d_2; \quad \alpha_\Sigma=\alpha_1\alpha_2\alpha_3$$

Here, for comparison, the transient processes in an extremal system without correcting link with respect to ψ_n have been plotted, in which $W_{xM}(s)$ and $W_{\mu u}(s)$ are the same as given above, and low-frequency filter with transfer function

$$W_\varphi(s)=\frac{1-e^{-sT}}{s}\,\frac{1}{\tau_3 s+1}$$

is included into the main extremal-control network z transform of $W_\varphi(s)$ is

$$W_\varphi^*(z)=\frac{1-d_3}{z-d_3}$$

where $d_3 = e^{T/\tau_3}$.

Bearing this remark in mind, for the given case, the equation of the dynamics (41) of the system adopts the form

$$A'_0e_{n+3}+A'_1e_{n+2}+A'_2e_{n+1}+A'_3e_n$$
$$-\alpha_\Sigma(1-d_1)[e^2_{n+2}-e^2_{n+1}+d_3(e^2_{n+1}-e^2_n)](-1)^n$$
$$=\Delta^3\psi_{n+1}+[(1-d_1)+(1-d_3)]\Delta^2\psi_{n+1}$$
$$+(1-d_1)(1-d_3)\Delta\psi_{n+1} \qquad (55)$$

where

$$A'_0=1; \quad A'_1=2a_M\alpha_\Sigma(1-d_1)-(1+d_1+d_3);$$
$$A'_2=2a_M\alpha_\Sigma(1-d_1)(1-d_3)+d_1+d_3+d_1d_3;$$
$$A'_3=-2a_M\alpha_\Sigma(1-d_1)d_3-d_1d_3; \quad \alpha_\Sigma=\alpha_1\alpha_2\alpha_3$$

As can be seen from the curves in *Figure 6*, an increase in β (the rate of drift of the extremum) leads to the loss of the stability of the system (55). Thus the introduction of compounding links with respect to ψ_n not only improves the quality of the system, but also preserves its stability, *thus extending the sphere of application of extremal systems to the case of high extremum drift rates.*

References

[1] SCHIPANOV, G. V. Theory and method of design of automatic controllers. *Automat. Telemech., Moscow* 1 (1939)

[2] KULEBAKIN, V. S. The theory of invariance of regulating and control systems. *Automatic and Remote Control.* p. 106. 1961. London; Butterworths

[3] PETROV, B. N. The invariance principle and the conditions for its application during the calculation in the design of linear and non-linear systems. *Automatic and Remote Control.* p. 117. 1961. London; Butterworths

[4] IVAKHENKO, O. G. *Automatika* (1961)

[5] KOSTYUK, O. M. *Automatika* 1 (1961)

[6] BELYA, K. K. The invariance of the controlled magnitude of an automatic device from certain of its parameters. *Izv. Akad. Nauk SSSR, Otdel Tekhn. Nauk, Energ. Automat.* 6 (1961)

[7] Invariance theory and its application in automatic devices. *Trud. Soveshch. Sostoyavshegosya v g. Kiev*, 16–20 sent., 1958 (Proc. of a meeting held in Kiev, Sept. 16–20, 1958), Moscow, 1959

[8] TSYPKIN, YA. Z. *Automatika* 1 (1958)

[9] TOU, J. Digital compensation for control and simulation. *Proc. Inst. Radio Engrs, N.Y.* Vol. 45, No. 9 (1957)

[10] KREMENTULO, YU. V. *Automatika* 1 (1962)

[11] KREMENTULO, YU. V. *Automatika* 2 (1960)

[12] KUNTSEVICH, V. M. *Automatika* 1 (1962)

[13] GRISHCHENKO, L. Z., and BOLDYREVA, D. F. The invariance of automatic sampled-data control systems. *Automatika* 2 (1962)

[14] STREITZ, V., and RUZHICHKA, I. The theory of autonomy and invariance of multiparameter control systems with digital controllers. *Izv. Akad. Nauk SSSR, Otdel Tekhn. Nauk. Energ. Automat.* 5 (1961)

[15] FEDOROV, S. M. Delay in the synthesis of servosystems with digital computers. *Izv. Akad. Nauk SSSR, Otdel Tekhn. Nauk Energ. Automat.* 4 (1961)

[16] TSYPKIN, YA. Z. *Teoriya Impulsnykh Sistem* (Theory of sampled-data systems) 1958. Moscow; Fizmatgiz

[17] BURSHTEIN, I. M. Solving equations of multiloop sampled-data systems. *Automat. Telemech., Moscow* 12 (1961)

[18] RAGAZZINI, J. R., and FRANKLIN, G. F. *Sampled-data Control Systems.* 1958. New York; McGraw-Hill

[19] JURY, E. J. *Sampled-data Control Systems.* [illegible]. New York;/ John Wiley. [illegible]. London; Chapman and Hall

[20] LENDARIS, G. G. and JURY, E. J. *Input-output Relationships for Multisampled-loop Systems Applications and Industry.* Jan. 1960

[21] TOU, J. A simplified technique for determination of output transforms of multiloop multisampler variable-rate discrete-data systems. *Proc. Inst. Radio Engrs, N. Y.* 49, 3 [illegible]

[22] TOU, J. *Digital and Sampled-data Control Systems.* 1959. New York; McGraw-Hill

[23] SALZER, G. M. Signal-flow reduction in sampled-data systems. *Wescon Conventional Record, Inst. Radio Engrs, N. Y.* Pt IV (1957)

[24] TOU, J. Statistical design of linear discrete-data control systems *via* the modified z-transform method. *J. Franklin Inst.* 271, 4 (1961)

[25] CHANG, S. S. L. Optimization of the adaptive function by the z-transform method. *A.I.E.C. Conf.* Pap. NCP 59–1296 (see also *Synthesis of Optimum Control Systems.* Ch. 10, 11. 1961. New York; McGraw-Hill)

[26] VAN-NEIS, R. I. *Automatika* 1, 2 (1961)

[27] IVAKHNENKO, A. G. Comparison of cybernetic extremal sampled-data systems characterized by extremum search strategy. *Automatika* 3 (1961)

[28] FELDBAUM, A. A. *Vychislitelnye Ustroistva v Avtomatika* (Computers and Automation) 1959. Moscow; [illegible]

[29] KUNTSEVICH, V. M. A study of sampled-data extremal systems with extremum drift. *Automat. Telemech., Moscow* 7 (1962)

[30] KUNTSEVICH, V. M. Invariance of sampled-data extremal systems without disturbance links. *Automatika* 3 (1962)

[31] TUNIK, A. A. *Automatika* 6 (1962)

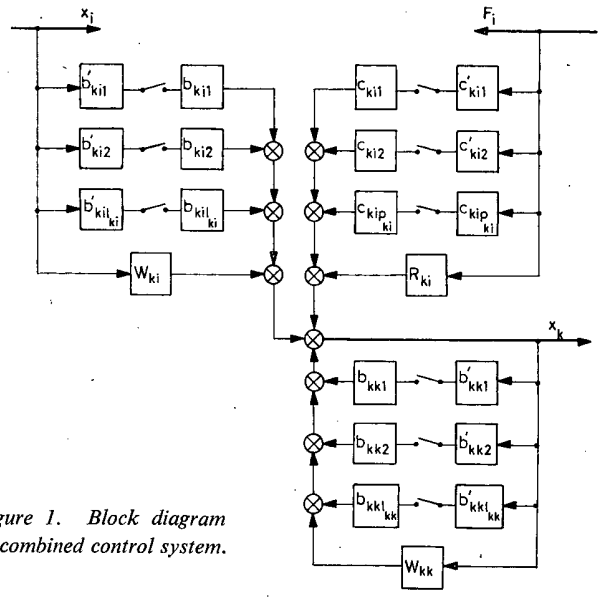


Figure 1. Block diagram of combined control system.

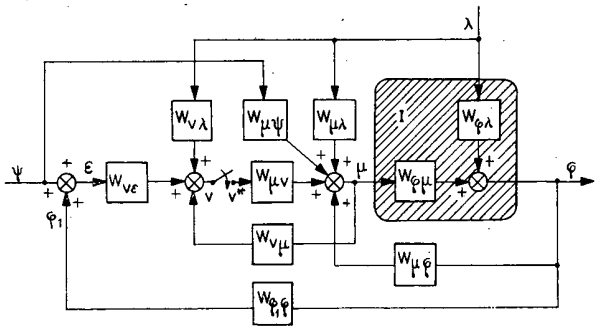


Figure 2. Block diagram of combined control system
I: plant

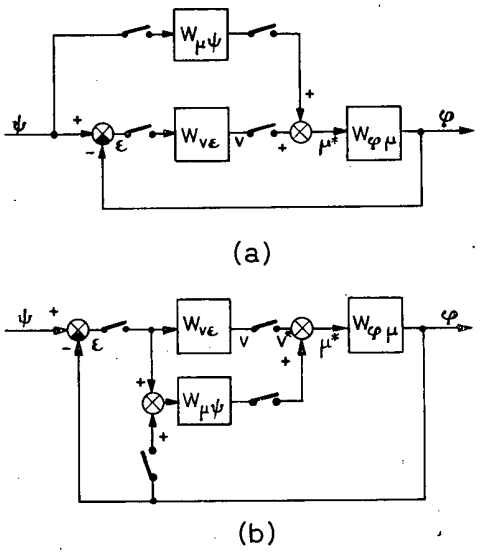


Figure 3. Block diagram of servosystems: (a) with direct link with respect to assignment; (b) with indirect link with respect to assignment

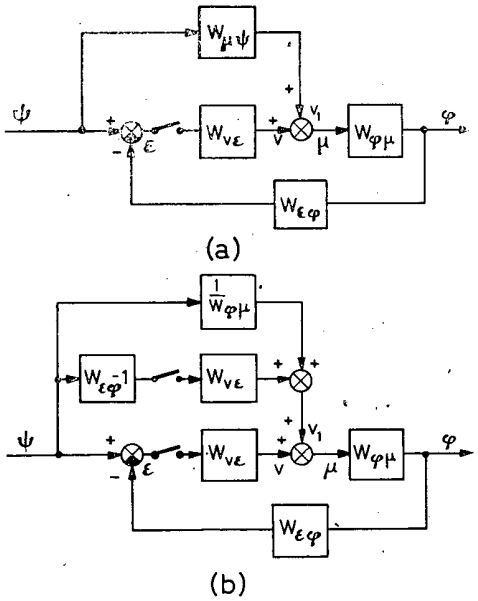


Figure 4. Block diagram Structural scheme of combined servosystem

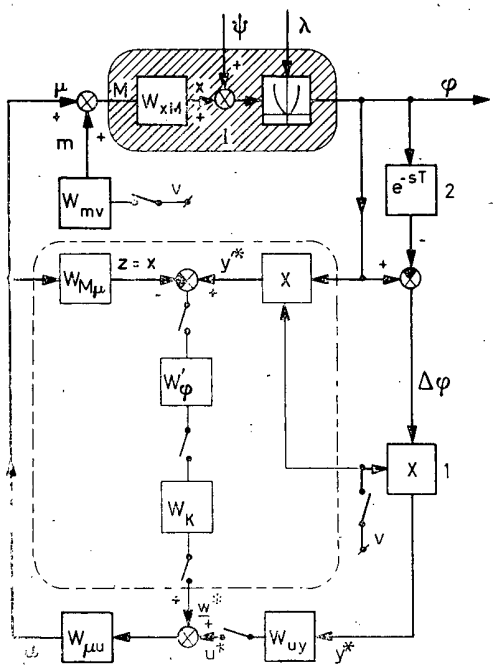


Figure 5. Block diagram of difference-type sampled-data extremal system with indirect compounding link
I: plant; 1: multiplying unit; 2: memory element

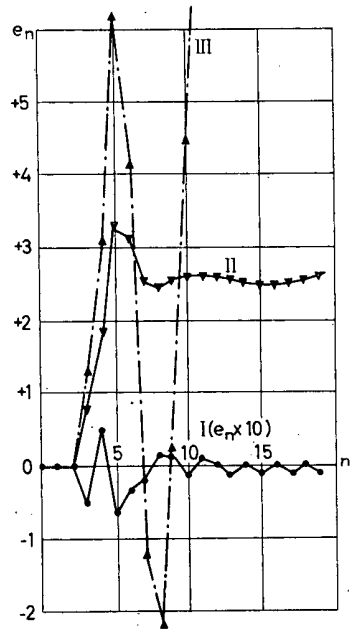


Figure 6. Transients of extremal system for $\psi_n = \beta n$, $\Delta\lambda_n = 0$
I: in system (54) with compounding link for satisfaction of condition (50); ($\alpha_1\alpha_2 = 0{\cdot}4$; $\alpha_3 = 1$; $d_1 = 0{\cdot}4$; $d_2 = 0{\cdot}8$; $\beta = 3{\cdot}5$);
II: in system (55) (ditto, but $d_1 = d_2 = 0{\cdot}4$; $\beta = 2$);
III: in system (55) (ditto, but for $\beta = 3{\cdot}5$)

Optimization and Invariance in Control Systems with Constant and Variable Structure

B. N. PETROV, G. M. ULANOV and S. V. EMELYANOV

Invariance and Optimization in Automatic Control Systems

Optimization of Automatic Control Systems and K (D) Image Theory

The object of the general theory of optimization of automatic control systems with respect to accuracy is the optimal synthesis of control systems operating under conditions of continuously-acting disturbances.

In the deterministic set-up of the problem[1–3, 7, 8] the optimality criterion is the achievement of the highest degree of accuracy of the automatic control system, as measured by the error ε, which is equal to the difference between the desired $g(t)$ and the realized $x(t)$ value of the state of the system $\varepsilon = g(t) - x(t)$. In the case of static synthesis the optimal system found from the probability characteristics of the controlling signal and the interference, has a transfer function Φ_{opt}, and possesses the greatest accuracy only in the mean.

The main results relating to the construction of optimal systems in the case of the deterministic set-up, have been obtained by the theory of invariance, on the basis of which there can be effected the construction of automatic control systems with an error ε, equal to zero or extremely small in the presence of disturbances, the measurement or use of which for the purposes of control is feasible. The conditions of the theory of invariance of automatic control systems, in the case when disturbance links do not nullify the numerator of the transfer function (and thus the corresponding transfer function), and when $f(t)$ is specified, are expressed with the aid of the $K(D)$ image introduced by Kulebakin

$$K(D) \cdot f(t) = 0, \quad K(D) \not\equiv 0, \quad f(t) \not\equiv 0 \ldots \qquad (1)$$

$K(D)$ and $f(t)$ are linked by the conditions of the operator $K(D)$ image of the functions[1]. In this case for a stable system its transfer function must either be the conform $K(D)$ image or have this operator $K(D)$ image as co-multiplier.

In the statistical set-up, with regard to determination of the transfer function of a control system in the case when it has an infinite memory, according to the mean-square error minimum criterion, one of the main results was obtained by Wiener. Obviously, in one case it is possible to establish precisely the correspondence of optimal systems in the case of the statistical and deterministic set-up of the problem. When the dispersion $f(t)$ tends to zero, Wiener's optimal system and the optimal system as determined by the conditions of invariance coincide and should, strictly speaking, lead to the same results. The generality of systems obtained in this case according to Wiener, and of invariant systems, in particular systems meeting the condition of Kulebakin's $K(D)$ image, are demonstrated. Taking the interval of observation of $f(t)$ to be infinite, and thus being concerned only with the forced output of the system, the transfer function of a Wiener optimal system is characterized by the magnitude of the *MS* error $\bar{\varepsilon}^2$ (ref. 6):

$$\bar{\varepsilon}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \{S_n(\omega) - |\Phi_{opt}(j\omega)|^2 S_f(\omega)\} \, d\omega \ldots \qquad (2)$$

$S_f(\omega)$ is the spectral density of $f(t)$, $S_n(\omega)$ the spectral density of the desired output signal. In the reviewed problems of control for stabilization $S_n(\omega)$ is conformally equal to zero, since, with complete filtration of external disturbance $f(t)$, the desired output of the system must be conformally equal to zero. The conditions of zeroth error $\bar{\varepsilon}^2_{min} = 0$ lead to the following requirement in respect of the optimal transfer function of an automatic control system:

$$\bar{\varepsilon}^2 = 0 \qquad S_n(\omega) = 0 \qquad (3)$$

$$|\Phi_{opt}(j\omega)|^2 \qquad S_f(\omega) = 0 \qquad (4)$$

The latter can be satisfied for $\Phi(p) \cdot f(t) = 0$, which is a sufficient condition.

In the case indicated, when

$$\Phi(p) = \frac{\Delta_1(p)}{\Delta(p)} = 0$$

where $\Delta_1(p)$ is the numerator of the transfer function, and $\Delta(p)$ is the characteristic polynomial of the automatic control system, expression (4) can be found for (*a*) $\Delta(p) = 0$ or (*b*) $K(p) \to \infty$, where $K(p)$ is the coefficient of transfer of the automatic control system (the characteristic equation of the control system is $\Delta(p) = K(p) + 1 = 0$).

The above-mentioned conditions correspond to the known conditions of invariance, the realization of which in physical systems is determined specially.

Without individually examining the above-mentioned possibilities (for $\Phi(p) = 0$), the case of the non-zero operator $\Phi(p) \neq 0$ will be considered.

If $\Phi_{opt} \not\equiv 0$ and $S_f \not\equiv 0$ the satisfaction of condition (4) is possible when

$$\Phi_{opt}(p) \cdot f(t) = 0 \qquad (5)$$

This requirement corresponds to the condition of invariance optimal according to Wiener in respect of disturbance $f(t)$, and coincides with the $K(D)$ image[1]. An analogous method is used to establish the community of invariant systems and systems optimal according to Wiener, in the case of other control problems. Thus the $K(D)$ image can serve as a tool for automatic control systems optimatization theory.

As an example, consideration is given to the forced motion of an automatic control system under the influence of an external disturbance, which is described by the equation

$$\Delta(p)\cdot x(t)=(p^2+\omega_K^2)\sin\omega_K t$$

The transfer function of system $\Phi(p) = p^2 + \omega_K{}^2/\Delta(p)$, by virtue of condition (5) corresponds to an optimal system, since it contains the $K(D)$ image of the action $f(t)$ as a comultiplier ($p^2 + \omega_k^2$ is the $K(D)$ image of $f(t) = \sin \omega_k t$).

Then, according to condition (4), the function $|\Phi(j\omega)|^2$ and $S_f(\omega)$ will respectively have the form of *Figure 1*.

The product of the function $|\Phi(j\omega)|^2 S_f(\omega)$ equals zero, since $|\Phi(j\omega)|^2 \geq 0$ when $\omega \neq \omega_K$, $|\Phi(j\omega)|^2 = 0^2$ when $\omega = \omega_K$

$$S_f(\omega)=\delta|\omega-\omega_K|\begin{cases}0 & \omega\neq\omega_K\\ \delta_{\text{funct}} & \omega=\omega_K\end{cases}$$

Generalization of $K(D)$ Image Theory for the Case of Statistically Given Disturbances $f(t)$

The $K(D)$ image theory expounded in the works of Kulebakin, was developed for the case of a disturbance $f(t)$, preset as a determined function of time t. To the class of functions particular, those which permit approximation of $f(t)$, as accurate as one likes, by integrals of linear differential equations, homogeneous and having constant coefficients. Shannon[9] has shown that a very broad class of functions, with the exception of hyper-transcendental functions and ξ functions, may also be approximated by the solutions of homogeneous differential equations with constant oefficients.

The need to develop statistical methods in the theory of invariance and in particular in the case of $K(D)$ images is explained by the following. The theory of invariance up to ε depends essentially upon the form of $f(t)$. The absolute invariance of automatic regulation and control systems in the case when the transfer function of the systems, as a function from $f(t)$ equals zero, is generally speaking real for any $f(t)$, constrained with respect to the modulus, in particular in relation to those about which information is missing.

In the case of the $K(D)$ image the effect of absolute invariance may only be observed for a completely defined function $f(t)$, knowledge of which, as a determined function of t, must be available with a probability of 1. Thus essential for the theory of invariance is knowledge about $f(t)$, which is nesessary in different cases with a probability from 0 to 1, particularly when investigating invariance with accuracy up to ε. In the case when $f(t)$ is given in a probabilistic sense, the effect of invariance—particularly from the viewpoint of the $K(D)$ image theory—was not examined, and the theory of invariance itself is not developed at the present time. An attempt is made below to apply the theory of statistical optimization to the determination of the statistical probabilistic conditions of automatic control systems invariance, and generalize the theory of $K(D)$ images for this case. Henceforward, as previously, we are examining the effect of invariance, the class of statistical actions $f(t)$ and control systems relating only to stationary systems and stationary actions $f(t)$.

Approximate Conditions of Optimalization Using the $K(D)$ Image in the Case when Dispersion is Present

In the well-known works of Kolmogorov[10] and others it is shown that any stationary random process may be represented as the limit of a sequence of processes with a discrete spectrum. The general expression of a stationary random process $f(t)$ in this case may be as follows:

$$f(t)=\sum_{K=1}^{n} a_K\sin(\omega_K t+\varphi_K) \qquad (6)$$

where $a_1, a_2, a_3, \ldots, a_K, \ldots, a_n$ are uncorrelated random magnitudes with mean value zero, i.e.,

$$M_{a_i}=0 \qquad i=1,2,\ldots,n$$

$$M_{a_i}M_{a_j}=0 \qquad i\neq j$$

where M is the sign of the mathematicel expectation.

It is also known[6, 10] that for each stationary process $f(t)$ it is possible to indicate a number ε as small as desired and as large as convenient an observation time range thereof T, for which there exist such pairwise uncorrelated random magnitudes $a_1, a_2, \ldots, a_n$ that the completeness of approximation to the series $\sum_{K=1}^{n} a_K \sin(\omega_K t + \varphi_K)$, determined by the mean-square difference, will be such that

$$M|x(t)-\sum_{K-1}^{n} a_K\sin(\omega_K t+\varphi_K)|^2\leq E$$

It has thus been shown that each stationary random process $f(t)$ can be approximated as accurately as desired by the sum of harmonic oscillations with random uncorrelated amplitude and phase. Most essential henceforward is the fact that ω_k characterizes the constant frequencies of process $f(t)$.

For the above series the correlational function $R_f(\tau)$ has, as it is known, the form

$$R_f(\tau)=\sum_{K-1}^{n}\frac{a_K}{2}\cos w_K t\,(M\{f(t)\}=0)$$

where ω_1 is the lower frequency of the spectrum of the random process, equal to $\omega_1 = 2\pi/\tau_{\max}$, $\tau_{\max}$ is an interval of time, beginning with which $|R_f(\tau) < \xi| R_f(0)$ where ξ is usually taken to equal 0·05.

For the $R_f(\tau)$ under consideration, the spectral density $S_f(\omega)$ represents a discontinuous function, consisting of δ functions of the form

$$S_f(\omega)=\sum_{K=1}^{n}\frac{a_K^2}{2}\delta(\omega-|\omega_K|) \qquad (7)$$

By virtue of the foregoing, the condition of an optimal control system is given by the expression

$$|\Phi_{\text{opt}}(j\omega)|^2 S_f(\omega)=0$$

or on the basis of (7)

$$|\Phi_{\text{opt}}(j\omega)|^2\cdot\sum_{K=1}^{n}\frac{a_K}{4}\delta(\omega-|\omega_K|)=0$$

Since the second co-multiplier of (7) characterizes the spectral density of some periodic function, the expression obtained may be written in the form

$$\Phi_{\text{opt}}(p)\cdot a_1\sin(\omega_1 t+\varphi_1)+\Phi_{\text{opt}}(p)\cdot a_2\sin(\omega_2 t+\varphi_2)+\ \ldots$$
$$+\Phi_{\text{opt}}(p)\,a_n\sin(\omega_n t+\varphi_n)=0$$

In this expression the magnitudes $a_1, a_2, \ldots, a_n$ and $\varphi_1, \varphi_2, \varphi_n$ are random, undetermined uncorrelated magnitudes, ω_K are constant for the given $f(t)$. For determination of Φ_{opt} the fact that a_K, φ_K are unknown is not essential, since $\Phi_{opt}(p)$, being the $K(D)$ image of $f(t) = \sum_{K=1}^{n} a_K \sin(\omega_K t + \varphi_K)$ is only determined by the frequency parameter ω_k. Since $\Phi_{opt}(p)$ for each partial frequency ω_K of the spectrum equals $p^2 + \omega_K^2$, the following will be the general expression of $\Phi_{opt}(p)$

$$\Phi_{opt}(p) = \left\{ \prod_{K=1}^{n} (p^2 + \omega_K^2) \right\} \Phi_0(p)$$

where $\Phi_0(p)$ is the remaining comultiplier of the function $\Phi_{opt}(p)$ after the removal from it of $\prod_{K=1}^{n} (p^2 + \omega_K^2)$.

The general problem of the approximate optimization $\Phi_{opt}(p)$ of a system in the presence of a random stationary disturbance $f(t)$ is thus solved with the assistance of the $K(D)$ image. Expansion of the stationary random process $f(t)$ into series (6) is a complex problem and it should be carried out on the basis of a preliminary examination of the process $f(t)$. So henceforward consideration is given to an assumed case in which the process $f(t)$ can be characterized by the presence of several main periodic oscillations in the spectrum. In this case the construction of systems satisfying the condition of the $K(D)$ image is facilitated by the limitation of n. In a number of practical examples of the use of the $K(D)$ image for dynamic systems of the damping type, the conditions of the $K(D)$ image are approximately satisfied only for one $n = 1$. The conditions of search of systems satisfying the requirements of $K(D)$ images may be effected on the basis of the statistical properties of $f(t)$. In the above case the automatic control system under consideration must satisfy the condition

$$K(D) \cdot \sum_{K=1}^{n} a_K \sin(\omega_K t + \varphi_K) = 0 \qquad (8)$$

Noting that the $K(D)$ image is itself invariant to random magnitudes of the series $f(t)$ to the random amplitude a_K and phase φ_K, and depends only on the determined values of ω_k, we shall find the $K(D)$ image for $\varphi_K = \pi/2$ and $a_K = a^2/2$ $(t - \tau)$.

Condition (8) will then have the form $K(D) \sum_{k=1}^{n} a_K/2 \cos \omega_K \tau = 0$ or $K(D) R(\tau) = 0$ where $R(\tau)$ is the correlation function of $f(t)$. Thus the condition of the invariance of the system to the disturbance $f(t)$, obtained on the basis of the theory of the $K(D)$ image, is equivalent to its invariance to the correlation function $R(\tau)$ of disturbance $f(t)$. The conclusion obtained is based on the expression of the stationary random process $f(t)$ (with a definite degree of accuracy) by a discrete Kolmogorov series[6], for which the corresponding spectral density is also the sum of discrete values in the form of δ functions. The possibility of using the discrete series (6) determines the applicability of the formula obtained for the case of an $f(t)$ given by continuous graphs of spectral density.

The condition of invariance to a random function, analogous to the condition derived above, can be obtained if the random function is expanded not into a Kolmogorov series, as was done above, but into a canonical series[5]*.

The random function $f(t)$ can be represented by its canonical expansion

$$f(t) = m_f(t) \sum_{v} V_v f^0(t)$$

where $m_f(t)$ is the mathematical expectation of $f(t)$, which will henceforward be put equal to zero: V_v are uncorrelated centred random magnitudes, coefficients of the canonical expansion, and $f_0(t)$ the coordinate functions of the canonical expansion.

The random coefficients V_v in the general form of canonical expansion of a random function are determined by the formula[5]

$$V_v = \overline{\Omega}^v F^0(t)$$

where $\Omega_{(v)}$ are arbitrary linear functionals, which must satisfy the conditions of biorthonogality for the mutual 'non-correlatedness' of the magnitude V_v; $f^0(t)$ is a centred random function $(F^0(t) = \Sigma_v V_v f^0_v)$.

The condition of invariance of a control system to disturbance $f(t)$ will be written in the form:

$$\Phi_{opt}(p) \cdot F^0(t) = 0 \quad (\Phi_{opt}(p) \neq 0) \qquad (9)$$

The coordinate functions $f(t)$ in the general form of canonical expansion of a random function are determined from the formula

$$f_v^0(t) = \frac{1}{D_v} \Omega_\tau^{(v)} R_f(\tau)$$

where D_v is the dispersion of an elementary random function, and $\Omega\tau^{(\psi)}$ is an arbitrary linear functional, the lower index of which signifies that this functional is applied to $R_f(t, \tau)$, viewed as a function τ at a fixed value of t.

Substituting into (9) the values of the coordinate functions and of coefficients V_v

$$\Phi_{opt}(D) \cdot \sum_{v} \overline{\Omega}^v f^0_{c_v}(t) \cdot \frac{1}{D_v} \Omega_c^{(v)} R_f(\tau) = 0 \qquad (10)$$

($\overline{\Omega}$ is a functional conjugate with Ω). The expression (10) is represented in the form

$$\sum_{v} \overline{\Omega}^{(v)} f_v(t) \cdot \frac{1}{D_v} \Phi_{opt}(p) R_f(\tau) = 0 \qquad (11)$$

For the identical equality of (11) to zero it is necessary and sufficient with $F^0(t) \neq 0$, $\Phi_{opt}(p) \neq 0$, $R_f(\tau) \neq 0$ that $\Phi_{opt}(p)$ be the $K(D)$ image of the correlation function $R(\tau)$ or contain it as a co-multiplier.

However, it should be noted that the representation of random processes by a spectral series (or canonical expansion) will practically always have a limited number of terms. This constraint causes the appearance of non-zero deflections on the output of the 'invariant' system (non-absolute invariance). The evaluation of this relation has its own significance and is not examined here.

Combined Tracking Systems with Variable Structure

Combined tracking systems are one of the most significant spheres of application of the principle of invariance in automatic control. In the combined system [*Figure 2 (a)*], reproduction of

* The idea of this solution belongs to A. S. Shatalov

the controlling action is implemented with the aid of a two-channel system or a system with two cycles: an open-loop cycle $\mu_2(p) = K_3(p)\,g(p)$ and a closed-loop cycle

$$x(p)=\frac{K_1(p)\,K_2(p)}{1+K_1(p)\,K_2(p)}\,g(p)$$

where $K_1(p)$, $K_2(p)$ are the transfer functions of the elements of the closed-loop cycle, $K_3(p)$ the transfer function of the open-loop cycle, μ_2 the output coordinate of the open-loop cycle and x the controlled coordinate.

The transient processes in such systems can be described by the linear non-homogeneous differential equation $M(p)\,\varepsilon = N(p)g(t)$ where $M(p)$ and $N(p)$ are operator polynomials relative to p, $p \equiv \mathrm{d}/\mathrm{d}t$, ε is the error signal. The independence of the error signal from the control action $g(t)$ is usually determined by the condition

$$N(p)=0 \tag{12}$$

in this case the forced component $\varepsilon_{\text{force}}(t)$ of the general solution of the equation of the system is conformally equal to zero. The links with respect to the controlling action $g(t)$ are selected in such a way as to satisfy condition (12). This is usually achieved by making the coefficients of the polynomial $N(p)$ consist of the difference of two magnitudes, one of which is determined by the disturbance effect (parameters of the open-loop cycle). It is practically impossible to satisfy condition (12) accurately.

An attempt will be made to solve this problem in another way. A tracking system will be constructed in such a way that the n-dimensional phase plane of a normal system of non-homogeneous differential equations, by which it is described relative to ε, where $\bar{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$, contains some $(n-1)$ dimensional hyperplane S, and it will be required that the motion of the state point in S be described by a system of homogeneous differential equations. Then, if the state point under any initial conditions and for any forms of $g(t)$ terminates its motion in this $(n-1)$ dimensional diversity of S, the error of signal ε will always tend to zero ($\varepsilon \to 0$) for any $g(t)$. In other words, the controlled coordinate $x(t)$ will reproduce any continuous $g(t)$ without static error, and the requirements for the operator $N(p)$, determined by condition (12), will be absent. If the function $g(t)$ has a discontinuity at some moments, then slight dynamic errors will appear at these moments. An attempt will be made to solve this problem, using the principles of construction of variable-structure automatic control systems[12].

Conditions of Invariance in Combined Tracking Systems with Variable Structure

In the domain, G, of an n dimensional space $\varepsilon_1, \ldots, \varepsilon_n$ let the motion of a dynamic system be described by a system of non-homogeneous differential equations with a discontinuous right-hand side

$$\frac{\mathrm{d}\bar{\varepsilon}}{\mathrm{d}t}=\bar{f}(\bar{\varepsilon},\bar{g}(t)) \tag{13}$$

Here

$$\bar{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_n),\ \bar{g}=(g_1,\ldots,g_m),\ \bar{f}=(f_1,\ldots,f_n)$$

$$f_i=\varepsilon_{i+1}\quad (i=1,2,\ldots,n-1)$$

$$f_n=-\sum_{i=1}^{n} a_i\varepsilon_i+\sum_{i=1}^{m}\psi_i(\bar{\varepsilon},\bar{g}(t)\,g_i(t)$$

where

$$\bar{\psi}_i(\bar{\varepsilon},\bar{g}(t))=\begin{cases} b_i \text{ for } \left(\sum_{j=1}^{n} c_j\varepsilon_j\right) g_i(t)>0\,\dagger \\ b_i^* \text{ for } \left(\sum_{j=1}^{n} c_j\varepsilon_j\right) g_i(t)<0 \end{cases}$$

a_i, b_i, b_i^*, c_j are constants, $g_i(t)$ is a function defined and continuous on the whole time interval t. Let the hyperplane S, set by the equation $\sum_{j=1}^{n} c_j\varepsilon_j = 0$ divide the domain G into sub-domains $G^+ (\sum_{j=1}^{n} c_j\varepsilon_j > 0)$ and $G^- (\sum_{j=1}^{n} c_j\varepsilon_j < 0)$, in which the vector function $\bar{f}(\bar{\varepsilon},\bar{g}(t))$ of system (13) is constrained and for any constant value of time t on the approach to S from G^+ and G^- there exist its limit values $\bar{f}^+(\bar{\varepsilon},\bar{g}(t))$ and $\bar{f}^-(\bar{\varepsilon},\bar{g}(t))$. On the approach of the solution $\bar{\varepsilon}(t)$ to some domain $U \subset S$ let the vector functions $\bar{f}^+$ and $\bar{f}^-$ be directed towards the hyperplane S ($f_N^- > 0$, $f_N^+ < 0$, where f_N^+ and f_N^- are the projections of the vectors $\bar{f}^+$ and $\bar{f}^-$ on to the normal to the hyperplane S, directed from G^- to G^+). Then, when $\bar{\varepsilon}(t)$ hits U there arises the so-called sliding mode and the solution of system (13) does not depend on $a_i, b_i, b_i^*, g_i(t)$. In fact in this case, as shown by Filippov[13], in the domain U^- there exists a solution $\bar{\varepsilon}(t)$ of system (13), and the vector $\mathrm{d}\bar{\varepsilon}/\mathrm{d}t = \bar{f}^0(\bar{\varepsilon},\bar{g}(t))$, where $\bar{f}^0 = (f_1^0, \ldots, f_n^0)$, lies in the hyperplane S and is determined by the values of the vector functions $\bar{f}^+$ and $\bar{f}^-$.

From the condition that $\bar{f}^0(\bar{\varepsilon},\bar{g}(t)) \in S$ there follows the linear relationship of the components of the vector $\bar{f}^0$

$$\sum_{j=1}^{n} c_j f_j^0 = 0 \tag{14}$$

where f_j^0 is the jth component of the vector $\bar{f}^0$ whence

$$f_n^0=\frac{-1}{c_n}\sum_{j=1}^{n-1} c_j f_j^0 \tag{15}$$

Hence the solution of system (13) for $\bar{\varepsilon}(t) \in U$ coincides with the solution of the system of similar homogeneous differential equations

$$\frac{\mathrm{d}\bar{\varepsilon}}{\mathrm{d}t}=\bar{f}^0(\bar{\varepsilon}) \tag{16}$$

Here

$$\bar{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_n)$$

$$f_j=\varepsilon_{j+1}\ (j=1,2,\ldots,n-1),\ f_n^0=\frac{1}{c_n}\sum_{j=1}^{n-1} c_j\varepsilon_{j+1}$$

c_j are constants.

Obviously the solution of system (16) does not depend on $a_i, b_i, b_i^*, g_i(t)$. Use will be made of this property of the solution of the system of non-homogeneous differential equations with a discontinuous right-hand side for the construction of a combined tracking system with variable structure.

† In the case $\left(\sum_{j=1}^{n} c_j\varepsilon_j\right) g_i(t) = 0$

$$\psi(\bar{\varepsilon},\bar{g}(t) = b_i \text{ for } \left(\sum_{j=1}^{n} c_j\varepsilon_j\right) g_i(t) \to +0$$

$$\psi_i(\bar{\varepsilon},\bar{g}(t)) = b_i^* \text{ for } \left(\sum_{j=1}^{n} c_j\varepsilon_j\right) g_i(t) \to -0$$

Let the structure, selected in a definite way, of the open-loop cycle of a combined tracking system [*Figure 3(b)*] change stepwise on some hyperplane $S = \sum_{j=1}^{n} c_j \varepsilon_j = 0$ in such a way that the movement of this servosystem is described by a system of non-homogeneous differential equations with a discontinuous right-hand side (13), where $\psi_i(\bar{\varepsilon}, \bar{g}(t)) = F[\Phi_i(\bar{\varepsilon}, \bar{g}(t))]$

$$\Phi_i(\bar{\varepsilon}, \bar{g}(t)) = \begin{cases} K_i \text{ for } \left(\sum_{j=1}^{n} c_j \varepsilon_j\right) g_i(t) > 0^{\dagger} \\ \qquad\qquad (i = 1, 2, \ldots, n) \\ K_i^* \text{ for } \left(\sum_{j=1}^{n} c_j \varepsilon_j\right) g_i(t) < 0 \end{cases}$$

K_i K_i^* are constants, determined by the open-loop cycle parameters. It is assumed (*a*) that the domain U exists, it includes the origin of the coordinates, and the solution of the system of differential equations (16) satisfies the given requirements on the quality of the process of control (control time and maximum dynamic error of the system must not exceed certain predetermined values;) (*b*) there exists a sufficiently large domain of initial conditions under which the solution of the system of equations (13) hits the domain U; (*c*) in the domain U there do not exist trajectories serving as sectors of limit cycles with a partially sliding regime.

Then the solution of the initial non-homogeneous system of differential equations (13) will depend on the controlling action $g_1(t)$ and the parameters of the closed-loop and open-loop cycles only up to the moment when $\bar{\varepsilon}(t)$ hits the domain U, where the solution coincides with the solution of the similar homogeneous system of differential equations (16) and depends only on the coefficients c_j. Thus in this case, in the reproduction of the controlling -actions $g_1(t)$ the magnitude $\varepsilon_1 \to 0$ on a finite interval of time t, and the controlled coordinate $x(t)$ reproduces $g_1(t)$ without static error. The quality of the process of control in such systems depends loosely on the variation of the parameters of the open-loop and closed-loop cycles, since the solution $\varepsilon(t)$ depends on these parameters only until it hits the domain U. It must be noted that in the systems under examination, the open-loop cycle for $g_1(t) \neq 0$ in isolated cases exerts an influence on the stability of the tracking system. In particular, as an example will demonstrate, even when the change of the parameters of the closed-loop cycle leads to the loss of the stability of the closed loop, then on the whole for $g_1(t) \neq 0$ the open-loop cycle with variable structure will ensure, in some domain of initial conditions, the stable operation of the tracking system. The above-listed properties of combined tracking systems with variable structure advantageously distinguish them from ordinary linear combined tracking systems.

An Example of a Combined Tracking System with Variable Structure

Let the equations of the individual components of a combined servosystem with variable structure have the form

$$\mu_1 = k_1 \varepsilon_1, \quad T_1 T_2 \ddot{g}_1(t) + (T_1 + T_2)\dot{g}_1(t) + g_1(t) = k_2 \mu_3$$

$$\Phi_1(\bar{\varepsilon}, g_1(t)) = \begin{cases} +K \text{ for } (c_1\varepsilon_1 + c_2\varepsilon_2)\, g_1(t) < 0^{*} \\ -K \text{ for } (c_1\varepsilon_1 + c_2\varepsilon_2)\, g_1(t) > 0 \end{cases}$$

where k_1, k_2, K, T_1, T_2, c_1, c_2 are constants. The block diagram of the system is depicted in *Figure 3(a)* and (*b*). In this case the combined tracking system, after the elimination of the intermediate coordinates μ_1, μ_2, μ_3, x is described by the following system of non-homogeneous differential equations with a discontinuous right-hand side:

$$\frac{d\bar{\varepsilon}}{dt} = \bar{f}(\bar{\varepsilon}, \bar{g}(t))$$

Here

$$\bar{\varepsilon} = (\varepsilon_1, \varepsilon_2), \bar{g} = g_1, g_2, g_3), f = (f_1, f_2) \tag{17}$$

$$f_1 = \varepsilon_2, \quad f_2 = -2b\bar{\varepsilon}_2 - \omega_0^2 \varepsilon_1 + g_3(t) + 2b g_2(t) + \psi_1(\bar{\varepsilon}, g_1(t))\, g_1(t)$$

where

$$2b = \frac{T_1 + T_2}{T_1 T_2} \qquad \omega_0^2 = \frac{1 + k_1 k_2}{T_1 T_2}$$

$$\psi_1(\varepsilon, g_1(t) = \frac{1 + \Phi_1(\bar{\varepsilon}, g_1(t))}{T_1 T_2}$$

$$g_2(t) = \frac{dg_1(t)}{dt} \qquad g_3(t) = \frac{dg_2(t)}{dt}$$

We shall examine the behaviour of a combined tracking system with variable open-loop structure which reproduces various controlling actions $g_1(t)$, while the parameters of the transfer function of the closed-loop cycle $K_2(p)$ can be chosen within wide limits.

The phase-plane method is used for analysis of the system. Let the controlling action $g_1(t) = A$, where A is a constant and the parameters of the tracking system k_1, k_2, T_1, T_2, K are selected in such a way as to satisfy the following conditions:

$$K \cdot k_2 > 1 \tag{18}$$

$$b^2 > \omega_0^2 \tag{19}$$

$$\frac{c_1}{c_2} = -b - \sqrt{(b^2 - \omega_0^2)^2} \tag{20}$$

Then for $g_1(t) > 0$ the phase plane of the system will have the form shown in *Figure 4(a)*, (*b*) and (*c*). In this case, under any initial conditions the state point will tend to hit the straight line $c_1\varepsilon_1 + c_2\varepsilon_2 = 0$ which serves as the boundary of discontinuity of the right-hand side of eqn (17) while on the boundary of discontinuity the vector functions $\bar{f}^+$ (sheet I) and $\bar{f}^-$ (sheet II) are always directed towards this straight line and, hence, when the state points hits it the solution of eqn (17) coincides with the solution of the similar homogeneous differential equation

$$\frac{d\bar{\varepsilon}}{dt} = f^0(\bar{\varepsilon}) \tag{21}$$

† In the case $\left(\sum_{j=1}^{n} c_j \varepsilon_j\right) g_i(t) = 0$

$$\Phi_i(\bar{\varepsilon}, \bar{g}(t)) = K_i \text{ for } \left(\sum_{j=1}^{n} c_j \varepsilon_j\right) g_i(t) \to +0$$

$$\Phi_i(\bar{\varepsilon}, \bar{g}(t)) = K_i^* \text{ for } \left(\sum_{j=1}^{n} c_j \varepsilon_j\right) g_i(t) \to -0$$

* For $(c_1\varepsilon_1 + c_2\varepsilon_2)\, g_1(t) = 0$

$$\Phi_i(\bar{\varepsilon}, g_1(t)) = +K \text{ for } (c_1\varepsilon_1 + c_2\varepsilon_2)\, g_1(t) \to +0$$

$$\Phi_i(\bar{\varepsilon}, g_1(t)) = -K \text{ for } (c_1\varepsilon_1 + c_2\varepsilon_2)\, g_1(t) \to -0$$

Here $$\varepsilon=(\varepsilon_1,\varepsilon_2),\ \overline{f}^0=(f_1^0,f_2^0)$$

$$f_1^0=\varepsilon_2,\ f_2=\frac{c_1}{c_2}\varepsilon_2$$

Thus the right-hand side of the equation determines the motion of the system only up to the moment when the state point hits the boundary of discontinuity, and then the motion of the system can be reflected by an equation without a right-hand side (21) or, after the appropriate transforms, by the equation

$$c_1\varepsilon_1+c_2\varepsilon_2=0 \qquad (22)$$

In this case, therefore, static error will be absent. We shall follow the variation of the static and dynamic properties of the system as the parameters of the transfer function of the closed-loop cycle $K_2(p)$ vary.

Let the parameters of the closed cycle vary in such a way that the closed loop of the system becomes unstable, e.g., consider that the sign is altered in front of the term $2b\varepsilon_2$ in eqn (17). In this case the system will also become unstable for any parameters of the linear transfer function of the open-loop cycle. When there is an open-loop cycle with a variable structure, the phase plane will have the form shown in *Figure 5*. As before, the state point, under any initial conditions, will hit the straight line (22), on which there exists a finite length *mn* which includes the origin of the coordinates 0, where the conditions of the existence of a sliding mode are satisfied. The tracking system will thus be stable. For a particular set of initial conditions the process will run without overshoot, and as before there will be no static error. Thus the variable-structure tracking systems under consideration are insensitive in relation to variation of the system parameters.

It is not difficult to show that for $g(t) < 0$ all the examined properties of the combined tracking system with variable structure will remain unchanged. We shall consider whether these properties of the system are preserved when reproducing other forms of controlling actions, e.g., $g_1(t) = \alpha t$, $Ae^{\alpha_1 t}$ where α, α_1 are constants.

In this case one will be dealing with a non-stationary phase plane. By examining the field of the tangents to the phase trajectories for various fixed moments of time t, the change of the directions of the vector functions ft and f can be followed and thus the answer given to the question of the existence of a section of sliding mode *mn* on the straight line (22) and the landing of the state point on this section.

Let the control action $g_1(t) = \alpha t$.

We shall examine the static and dynamic properties of a tracking system for the first case of combination of closed-loop cycle parameters. For the instant $t = 0$ [*Figure 6(a)*] the direction of the vector functions $\bar{f}^+$ and $\bar{f}^-$ in the vicinity of the straight line (22) is such that the section of sliding mode *mn* on straight line (22) is everywhere absent. However, with the time, beginning with some $t = t_1$, the field of the tangents to the phase trajectories changes in such a way that the vector functions $\bar{f}^+$ and $\bar{f}^-$ in the vicinity of the straight line (22) are everywhere directed towards this straight line [*Figure 6(b)*]. Since, with time [*Figures 8(c)*, *9(c)* and *10(c)*] the inclination of the tangents to the phase trajectories is deformed in such a way that at the limit it tends towards straight lines [*Figure 6(c)*], the above-mentioned static and dynamic properties when the system will also be preserved when the system is reproducing a controllin action of the kind underg review. Let the controlling effect be $g_1(t) = Ae^{\alpha_1 t}$. From the analysis of the variation of the fields of the tangents for various instants t, it follows that even when reproducing a transcendental controlling action, static error is absent.

With the aid of an electronic simulator we shall study the behaviour of such a combined tracking system with a variable structure in the reproduction of controlling actions of the form

$$g_1(t)=A,\alpha t+A_1,A_2e^{\alpha_1 t} \text{ where } A,\alpha,A_1,\alpha_1,A_2$$

are constants.

Let the parameters of the tracking system equal

$$T_1=1,\ T_2=1,\ k=1,\ k_2=1,\ K=2$$

As follows from the oscillograms in *Figure 8(a)*, *(b)* and *(c)* all the controlling actions under review are reproduced without static errors with a good quality of the transient processes.

We shall change any one of the parameters of the controlled plant, e.g., k_2 from the value $k_2 = 1$ to $k_2 = 10$, and follow the change of the static and dynamic errors of a combined tracking system with variable structure. As can be seen from the oscillograms in *Figure 8(d)*, *(e)* and *(f)* the static properties of the system have been preserved in this; as before, it reproduces, without static errors, all the types of controlling actions under consideration, while the dynamic properties have not suffered any qualitative changes—the time of the transient processes has been slightly reduced.

Conclusions

The paper considers the invariance of automatic control systems in the presence of statistically given disturbances. The invariance conditions, obtained on the basis of the $K(D)$ image theory, have been generalized for the case of statistically given disturbances. For stationary systems of automatic control and stationary disturbances $f(t)$ the conditions of the $K(D)$ images in relation to the disturbance prove to be equivalent to the condition of the $K(D)$ image in relation to its correction function.

A new principle has been proposed for the design of invariant tracking systems in relation to continuous functions of the controlling action, which ensure the absence of static error. It is shown when using an open-loop cycle with variable structure that it is possible to reproduce, without static errors, an extensive class of controlling-action functions. When selecting the open-loop cycle transfer function there is no need to satisfy the classical conditions of invariance, which require the right-hand side of the non-homogeneous differential equation to vanish. This property of the systems under consideration makes it possible to build invariant tracking systems without differentiation of controlling action. The variable-structure combined tracking systems considered are insensitive to the variation of the system parameters within a certain range.

References

[1] KULEBAKIN, V. S. *Uspekh. Mat. Nauk.* 6, No. 5 (1951), 211; *Dokl. Akad. Nauk.* 68, No. 5 (1949); *Dokl. Akad. Nauk.* 77, No. 2 (1951)

[2] PETROV, B. N., and ULANOV, G. M. Some problems of the theory of combined automatic control systems. *Trud. Sessii Akad. Nauk S.S.S.R.*, No. 5 (1956)

[3] PETROV, B. N. The principle of invariance and the condition of its use in designing linear and non-linear systems. *Automatic and Remote Control. Proc.* 1. 1960. London; Butterworths

[4] WIENER, N. *Extrapolation, Interpolation and Smoothing of Stationary Time Series.* 1949. New York

[5] PUGACHEV, V. S. *The Theory of Random Functions.* 1957. Moscow; State Publishing House of Physico-Mathematical Literature

[6] SOLODOVNIKOV, V. V. *The Statistical Dynamics of Linear Automatic Control Systems.* 1960. Moscow; State Publishing House of Physico-Mathematical Literature

[7] IVAKHNENKO, A. G. Electro-automation. *Izd. Akad. Nauk Ukr. S.S.R.* Pt. 1, II (1957)

[8] ULANOV, G.M. *Disturbance Control.* 1960. Moscow; Gosenergoizdat

[9] SHANNON, [redacted]. *J. Math. Phys. U.S.A.* 20 (1941)

[10] KOLMOGOROV, A. N. Stationary sequences in a Gilbert space. *MGU.* 2, No. 6 (1941)

[11] KAZAKOV, I. B. Approximate probability analysis of the accuracy of operation of essentially non-linear systems. *Avtomatika i telemekhanika* XVII, No. 5 (1956)

[12] EMELYANOV, S. V. The use of 'key' type non-linear correcting devices to improve the quality of second-order automatic control systems. *Avtomatika i telemekhanika* XX, No. 7 (1959)

[13] FILIPPOV, A. F. Differential equations with a discontinuous right-hand side. *Matem. sbornik* No. 1 (1960)

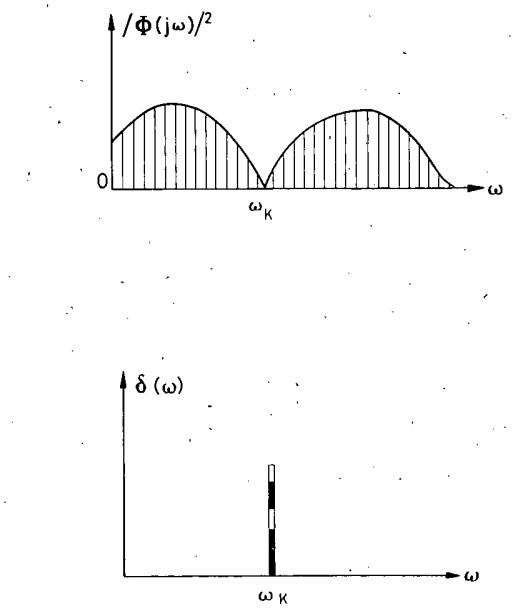
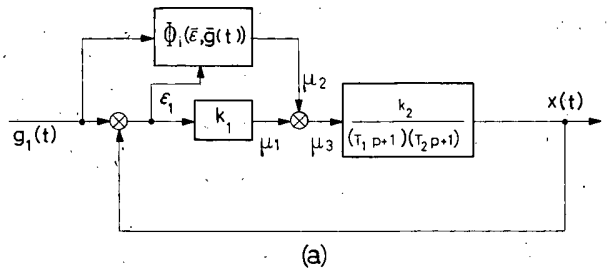
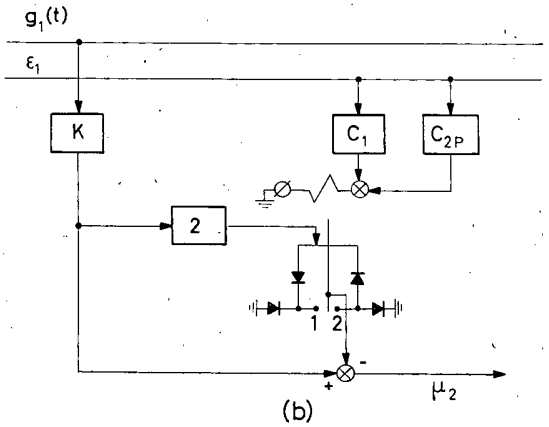
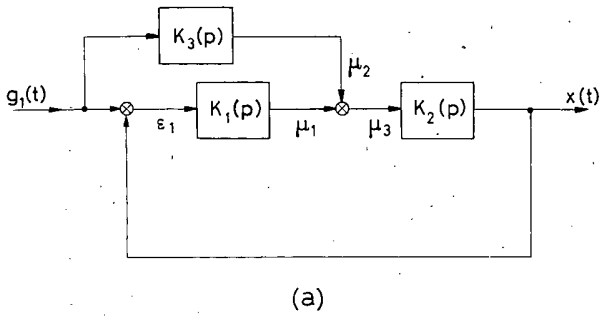


Figure 1







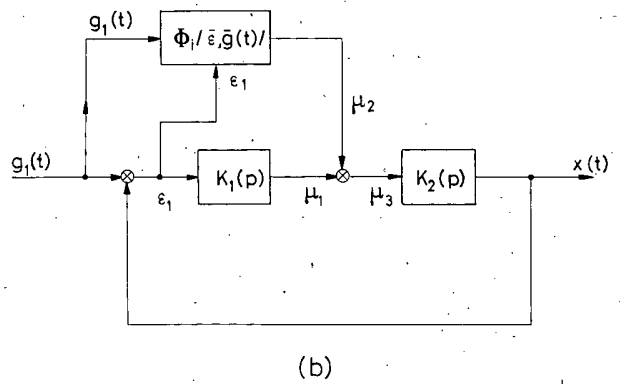


Figure 2

Figure 3

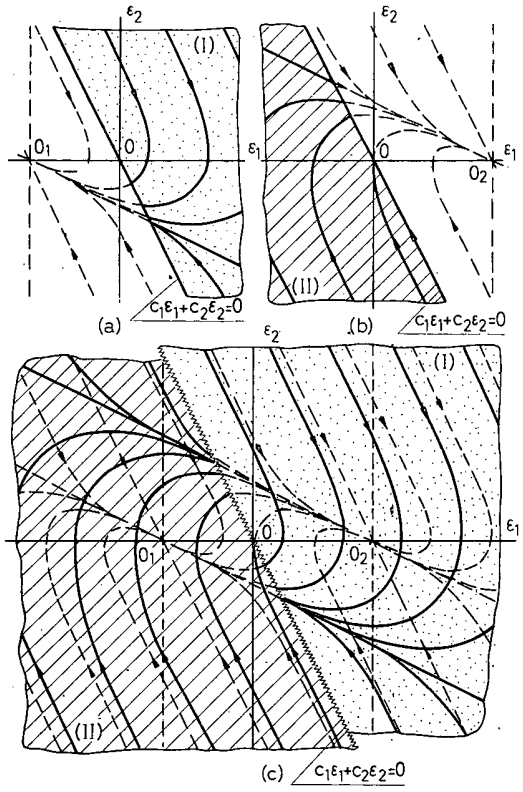


Figure 4

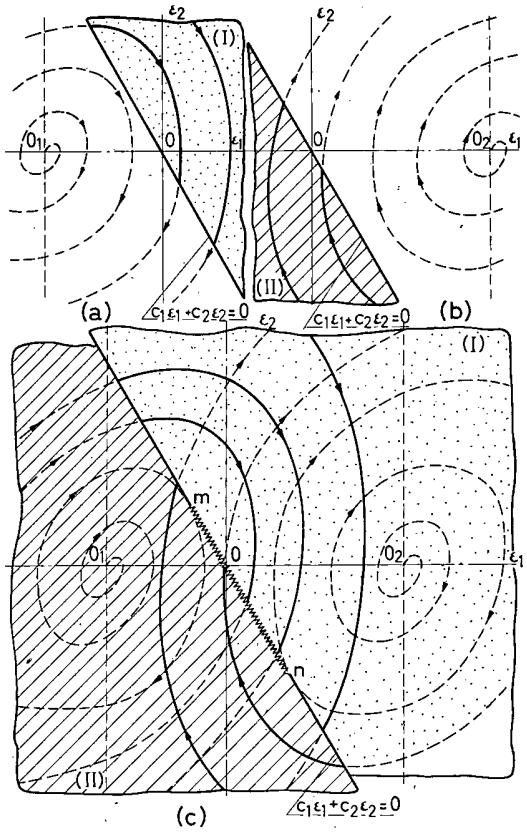


Figure 5

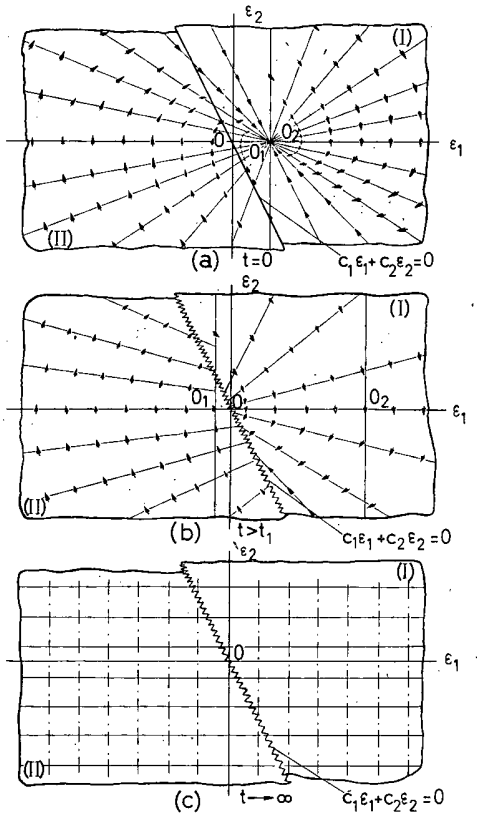


Figure 6

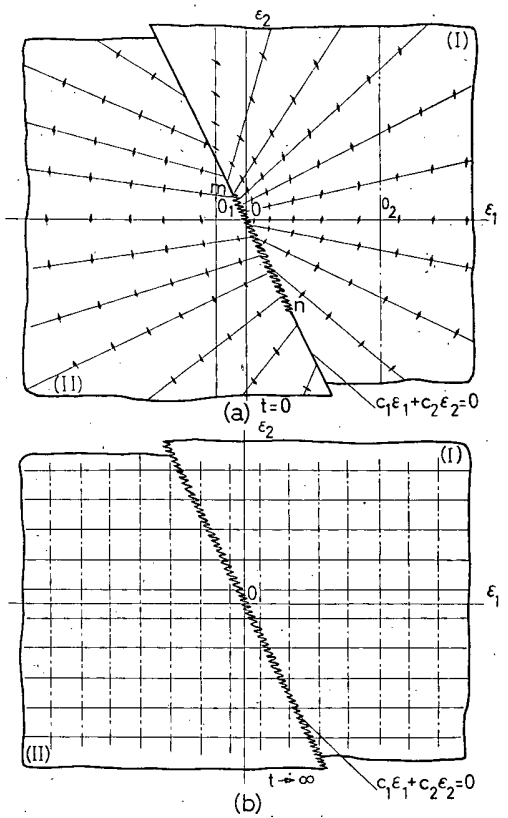


Figure 7

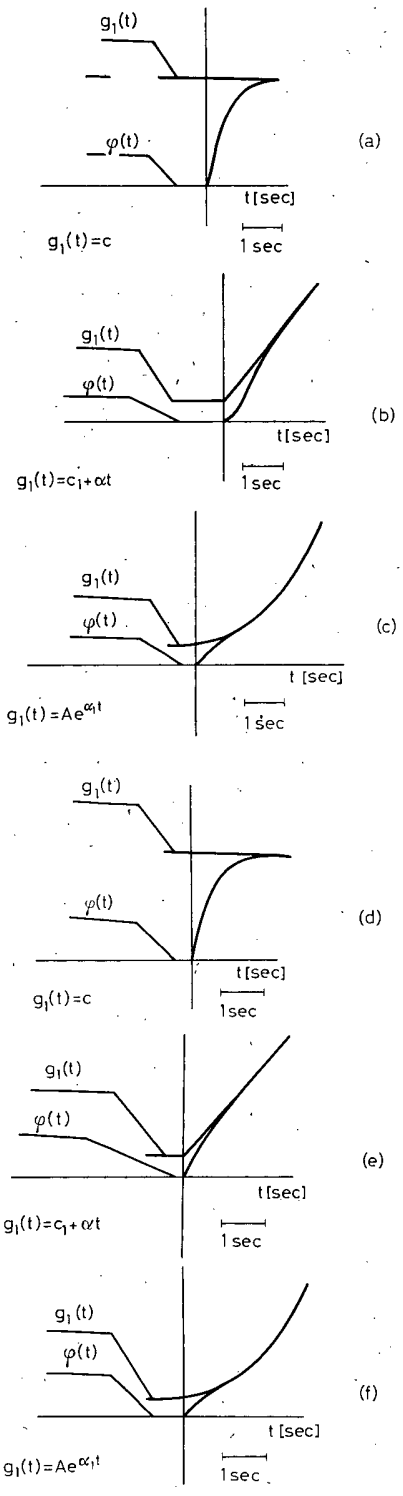


Figure 8

Time-optimal Systems with Random Noise Disturbances

V.V. NOVOSELTSEV

Introduction

This paper examines the problem of optimal control of a plant with constant coefficients, having one input and one output:

$$\dot{x} = f^1(x) + v \cdot f^2(x) \tag{1}$$

Here x is an n-dimensional vector which defines the state of the plant, v is the control signal sent to its input.

Functions $f^1(x)$ and $f^2(x)$ are defined and continuous for all x and continuously differentiable with respect to all coordinates of vector x:

$$x_i = \mathrm{d}^i x / \mathrm{d}t^i$$

Equation (1) is linear with respect to v and non-linear with respect to x, and is therefore somewhat more general than the equation

$$\dot{x} = Ax + Bu$$

usually considered for the case of a scalar control signal u.

Figure 1 shows a block diagram of the system under consideration in the presence of interference; the following symbols are used:

A—controller
B—controlled plant (1)
H—inertia-less plant coordinate metering channel
G—inertia-less control signal-to-plant channel
Z—master-signal channel
h, g and *z*—set random interference in channels *H, G* and *Z* respectively
u—control signal (scalar)
v—noise-distorted control signal
x_d—true state of plant
x_n—observed state of plant
x_z—set point of system phase space
x—vector of error $x = x_z - x_d$

The true-state point x_d has to be shifted into some small vicinity of the set point (origin of the coordinates). Then in the optimal system the following equality must be satisfied[1, 2]:

$$E\{T[x^{(0)}]\} = \min$$

The minimal value of $E\{T(x)\}$ will be denoted by $T^*(x)$. Then in the optimal system

$$E\{T[x^{(0)}]\} = T^*[x^{(0)}] \tag{2}$$

Here $x^{(0)}$ is the initial value of the error vector.

However, considerable difficulties are involved in calculating the system directly from this criterion. It is far simpler to use the criterion of the minimum time of the transient process for the mathematical expectation (the minimum time required to bring the system to a state of statistical non-displacement)[3, 4]. Usually considered for the determination of this time is the relationship $\xi = E\{x\}$, which describes the transient processes in some equivalent system without interference. A system in which is provided the minimum time of the transient process for the mathematical expectation $T(\xi) = \theta(x)$

$$\theta(x) = \min \equiv \theta^*(x) \tag{3}$$

will be termed optimal with respect to the criterion θ^*. A system in which condition (2) is satisfied, will be termed optimal with respect to the criterion T^*.

Functions $T^*(x)$ and $\theta^*(x)$ are defined and continuous for all points of the phase space and $T^*(x) \geq 0$, $\theta^*(x) \geq 0$ while

$$T^*(x) = \theta^*(x) = 0$$

when, and only when, the point x lies in the set vicinity of a finite point

$$\sum_{i=0}^{n-1} (x_i)^2 \leq \delta^2 \tag{4}$$

Consideration will be given to the state of the control system only at discrete moments of time, as in solving similar problems by the dynamic programming method. For this the small interval of time Δ will be introduced and it will be considered that on this interval the values of the control action and the interference signals remain invariant, but at moments of time $t = k\Delta$ they change stepwise. Interference on neighbouring intervals will be considered independent.

Control action u is constrained with respect to the modulus

$$|u| \leq N \tag{5}$$

To simplify the examination it will be assumed that both u and v are quantized in level with a sufficiently small pitch of quantization, and:

$$u \in \Omega(u) \qquad \Omega(u) = \{u_1, u_2, \ldots, u_r\}\ r < \infty$$

$$v \in \Omega(v) \qquad \Omega(v) = \{v_1, v_2, \ldots, v_l\}\ l < \infty$$

In such a case it is convenient to describe the influence of random interferences in the following way.

Since at the controller A there arrives only the value of the error x distorted by the noises along channels H and Z, instead of the correct, necessary control action u at each moment of time another control action, generally speaking distinct from

$u_0(x^{(k)})$, will be chosen. The probability of the choice of control action u at a given moment of time, when in fact control action u^0, denoted by $p_u{}^0{}_u$ is optimal, depends in the general case on the position of the image point x in the system phase space. Thus, in a relay system this probability is heavily influenced by distance of the image point from the switching surface. When this distance is great, the probability of error in choice of the control action is small, but as this distance is reduced even very weak interference can lead to error in the choice of control action. If one denotes the probability of an event which consists in the appearance on the output of A of signal u_m, whereas the optimal choice would be $u^0(x) = u_l$ by p_{ml}, then

$$\|p_{u^0u}(x)\| = \begin{Vmatrix} p_{11}(x)\, p_{12}(x) \dots p_{1r}(x) \\ p_{21}(x)\, p_{22}(x) \dots p_{2r}(x) \\ \dots\dots\dots\dots\dots \\ p_{r1}(x)\, p_{r2}(x) \dots p_{rr}(x) \end{Vmatrix} \tag{6}$$

The control action $u(x^{(k)}) = u_m$ chosen at the kth moment by controller A reaches the plant along the channel with noisy G, where, under the influence of interference g control action u becomes control action V. The probability of the transformation of u_i into v_j will be denoted by q_{ij}. Then

$$\|q_{uv}\| = \begin{Vmatrix} q_{11} q_{12} \cdots q_{1l} \\ q_{21} q_{22} \cdots q_{2l} \\ \dots\dots\dots \\ q_{r1}\, q_{r2} \cdots q_{rl} \end{Vmatrix} \tag{7}$$

Both matrices $\|p_{u^0u}{}^{(x)}\|$ and $\|q_{uv}\|$ can be joined into one, which will fully describe the action of all the interference upon the system:

$$\|p_{u^0v}\| = \|p_{u^0u}^{(x)} \cdot \|q_{uv}\| \tag{8}$$

Matrix (8) determines for each point of the phase space the probability of the arrival at the input of the plant of control action v_j, when u_i is the optimal action.

Basic Relationship for Time-optimal Systems with Interference

In the optimization of control systems with respect to the criterion T^* by the dynamic programming method, the following equation can be obtained:

$$T^*[x^{(k)}] = \Delta + \min_{u^{(k)} \in \Omega(u)} T^*[x^{(k+1)}] \tag{9}$$

Here $x^{(k)}$ and $u^{(k)}$ denote the values of x and u on the kth interval of time. Equation (9) is the basic relationship in solving such problems[2, 5]. This relationship will be given another form more suitable for the purposes of this paper[6].

In open form eqn (9) is written as follows:

$$T^*[x^{(k)}] = \Delta + \min_{(m)} \left\{ \sum_{j=1}^{l} p_{mj}(x^{(k)}) \cdot T^*[x^{(k+1)}]_j \right\} \tag{10}$$

$$m = 1, 2, \dots, r$$

In the latter equation $\min\{\alpha_m\}$ denotes the minimum of the numbers α_m, and $[x^{(k-1)}]$ the position of the image point at the $(k+1)$th moment of time, provided that at the kth moment the point was in the position $x^{(k)}$ and at the plant input there arrived control action v_j. Thus the mth term of the expression in brackets is simply the mathematical expectation of the time of the transient process in the choice at point $x^{(k)}$ of control action u_m. Averaging is performed for all the states of the system at the $(k+1)$th moment of time. The probabilities p_{mj} in (10) are elements of the matrix $\|p_{u^0v}\|$, which is determined by (6).

By introducing the sampling interval Δ eqn (1) can be rewritten in the form

$$x_i^{(k+1)} = x_i^{(k)} + \varphi_i^1 + v \cdot \varphi_i^2 \quad i = 0, 1, \dots, n-1$$

$$x_n^{(k+1)} = v^{(k+1)}$$

where, for brevity, is written $\Delta \cdot f^1 = \varphi^1$; $\Delta \cdot f^2 = \varphi^2$. In subsequent operations the relationship to $x^{(k)}$, where possible, is dropped. If in the expansion of T^* into a series, it is possible to limit ourselves (for sufficiently small Δ) to terms of no higher than the first order of smallness, then

$$T^*[x^{(k+1)}]_j \equiv T^*[x_0^{(k+1)}, x_1^{(k+1)}, \dots, x_{n-1}^{(k+1)}]_j$$
$$= T^*[x_0^{(k)}, x_1^{(k)}, \dots, x_{n-1}^{(k)}]$$
$$+ \sum_{i=0}^{n-1} \frac{\partial T^*[x^{(k)}]}{\partial x_i} [\varphi_i^1 + v_j \varphi_i^2] \tag{11}$$

Substituting (11) into each bracketed term in (10), after elementary transforms

$$\sum_{j=1}^{l} p_{mj} T^*[x^{(k+1)}]_j = T^* + \sum_{i=0}^{n-1} \varphi_i^1 \frac{dT^*}{\partial x_i}$$
$$+ \sum_{i=0}^{n-1} \left\{ \frac{dT^*}{\partial x_i} \cdot \varphi_i^2 \cdot \sum_{j=1}^{l} p_{mj} v_j \right\} \tag{12}$$

is obtained.

The following notations will be introduced

$$v_m^*(x^{(k)}) \equiv E\{v | u_m(x^{(k)})\} = \sum_{j=1}^{l} p_{mj}(x^{(k)}) \cdot v_j \tag{13}$$

$$[\xi^{(k+1)}]_m = E\{x^{(k+1)} | x^{(k)}, v_m^*(x^{(k)})\} \tag{14}$$

Here $v_m{}^*(x^{(k)})$ is the mathematical expectation of the signal v, and $[\xi^{(k-1)}]_m$ the mathematical expectation of the random vector $x^{(k+1)}$ for the selection at the point $x^{(k)}$ of control action $u^{(k)} = u_m$. Then, on the basis of (11)–(14), the initial relation (9) can be written in the form:

$$T^*[x^{(k)}] = \Delta + \min_{(m)} \{T^*[\xi^{(k+1)}]_m\} \quad m = 1, 2, \dots, r \tag{15}$$

Control action u_m, with which the minimum is reached, is the T^*-optimal control action at the point $x^{(k)}$ of the system phase space.

Equation (15) is the basic relation in examination of time-optimal systems with noise present.

The optimal control action at each moment of time must be so selected that the magnitude of $T^*[\xi^{(k+1)}]$ at the next step is minimal.

Control Algorithm in a System Optimal with Respect to Criterion T^*

Consideration is given to the point $x^{(k)}$ and the sequence of control actions $u^{(k)} = u_\alpha$, $u^{(k+1)} = u_\beta$, ..., $u^{(k+S)} = u_\sigma$ under the influence of which the image point shifts consecutively from position $x^{(k)}$ to positions $[\xi^{(k+1)}]_\alpha$, $[\xi^{(k+2)}]_{\alpha\beta}$, ..., $[\xi^{(k+S)}]_{\alpha\beta}$, ..., σ. Let the sequence $u_\alpha, u_\beta, \ldots, u_\sigma$ be selected in such a way that:

(*A*) The point $[\xi^{(k+S)}]_{\alpha\beta\cdots\sigma}$ lies in the vicinity of the set point (4).

(*B*) For all sequences $u^{(k)}$, $u^{(k+1)}$, ... $u^{(k+d)}$ where $d < S$, condition *A* is not satisfied.

Then the sequence $x_\alpha, u_\beta, \ldots, u_\sigma$ is optimal with respect to the criterion θ^*, and S is the number of steps in the optimal transient process with respect to the criterion θ^*. It will be temporarily assumed that there exists and is known a control algorithm optimal with respect to the criterion θ^*, which permits the construction of such a sequence of control actions for any x.

The recurrent relation (15) will now be applied to the point $[\xi^{(k+1)}]_m$:

$$T^*[\xi^{(k+1)}]_m = \Delta + \min_{(n)} \{T^*[\xi^{(k+2)}]_{mn}\} \quad n = 1, 2, \ldots, r \tag{16}$$

Substituting the resultant expression back into (15), one obtains:

$$T^*[x^{(k)}] = 2\Delta + \min_{(m,\,n)} \{T^*[\xi^{(k+2)}]_m\} \tag{17}$$

$$m = 1, 2, \ldots, r;\ n = 1, 2, \ldots, r$$

similarly

$$T^*[x^{(k)}] = s\cdot\Delta + \min_{\underbrace{(m,\,n,\,\cdots,\,l)}_{s}} \{T^*[\xi^{(k+s)}]_{mn,\,\cdots,\,l}\} \tag{18}$$

$$m = 1, 2, \ldots, r;\ n = 1, 2, \ldots, r;\ l = 1, 2, \ldots, r$$

One now selects $m = \alpha, n = \beta, \ldots, l = \sigma$. Then, by virtue of condition *A*, the second addenda in the right-hand side of (18) vanishes, and, bearing in mind *B*,

$$T^*[x^{(k)}] = S\cdot\Delta \tag{19}$$

can be written.

Thus in the case under consideration the duration of the transient process with respect to criteria T^* and θ^* is identical, and at each point of the phase space the control actions optimal with respect to θ^* and T^* coincide.

The control algorithm optimal with respect to criterion θ^*, ensures the optimality of the system with respect to criterion T^* as well.

Consideration is now given to the construction of algorithms optimal with respect to the criterion N. The plant studied is linear and is described by the equation

$$x_i = \sum_{j=0}^{n-1} a_j x_j \qquad i = 0, 1, \ldots, n-1$$
$$x_n = v \tag{20}$$

It will also be assumed that the control algorithm, which ensures time-optimality in the absence of interference, is known and set in the form of a switching surface in an n-dimensional space.

$$\psi(N, x_0, x_1, \ldots, x_{n-1}) = 0 \tag{21}$$

Equation (21) contains in an explicit form the magnitude N from eqn (5). In the absence of interference the optimal equation has the form

$$u^0 = N \operatorname{sign} \psi(N, x) \tag{22}$$

and only the values of $v = u = \pm N$ reach the plant input.

If on the control system (20)–(21) there are noises, then in place of a system with interference, an equivalent system without interference can be considered, in which instead of the coordinates $x(t)$ the relationships $\xi(t) = E\{x(t)\}$ are considered. The optimal control action in such an interference-free system will be optimal with respect to the criterion θ^* in the initial system with interference, and will hence be optimal with respect to the criterion T^*.

The maximum and minimum values of the signal v^* which reaches the plant input when $|u| \leq N$ will be denoted by v_M^* and v_m^*.

For the symmetrical matrices (6)–(8) $v_M^* = -v_m^*$. For simplicity, the examination will be confined to the case when the signals v_M^* and v_m^* are obtained following the selection on the controller of control actions $u = +N$ and $u = -N$ respectively.

Introduced here is the coefficient of efficiency of control in a system with interference, which has an obvious sense:

$$\gamma(x) = \frac{v_M^*(x)}{N} = -\frac{v_m^*(x)}{N} \tag{23}$$

If function $\gamma(x)$ is defined for all x, continuous and continuously differentiable with respect to x, then it is convenient to examine as an equivalent system a system for control of an equivalent non-linear plant. Eqn (20) is then written in the form:

$$x_i = \sum_{j=0}^{n-1} a_j x_j;\ x_n = \gamma(x)\cdot u \tag{24}$$

$$i = 0, 1, \ldots, n-1$$

and the constraint $|u| \leq N$ is retained.

The optimal control of plant (24), with the constraint $|u| \leq N$, conforms to the maximum principle[7].

It is noted that replacement of eqn (20) by eqn (24) is not obligatory. Using the results obtained in Fillipovs' work[8], it is also possible to examine eqn (21) directly, but with the replacement of constraint (5) by a constraint of more general form: $u \in Q(x)$.

The following three cases of the action of interference upon the system are considered:

(1) In the system, interference is present only in channel G. Here, as follows from (7), $\gamma = \text{const}$, and constraint (5) is replaced by the constraint $|u| \leq \gamma N$.

The plant equation remains unchanged. The optimal switching surface has the form

$$\psi(\gamma N, x_0, x_1, \ldots, x_{n-1}) = 0 \tag{25}$$

(2) Interference is absent from the channel G, but the influence of interference h and z manifests itself in the appearance of additive noise along the coordinates $x_0, x_1, \ldots, x_{n-1}$; at the controller A there arrive the values

$$x_0^* = x_0 + eq\,\eta_0$$
$$x_1^* = x_1 + eq\,\eta_1$$
$$\cdots$$
$$x_{n-1}^* = x_{n-1} + \eta_{n-1}$$

while the random component is constrained with respect to the modulus:

$$|\eta_i| \leq \eta_i^*, \qquad i = 0, 1, \ldots, n-1 \tag{26}$$

In this case it can be shown that the optimal switching line in a second-order control system has the form:

$$\psi(N, x_0 + \eta_0^* \operatorname{sign} x_0, x_1 + \eta_1^* \operatorname{sign} x_1) = 0 \tag{27}$$

Optimality is ensured for all points of the phase plane sufficiently remote from the set point. It may be assumed that an equation analogous to eqn (27) is also valid for a system controlling plants of a higher order with real roots.

(3) In the system there is present both interference in the channel G and unconstrained interference η.

In this case it is possible to construct approximately optimal control systems, in which the duration of the transient processes exceeds the minimum possible time by not more than the preset ε.

One such system is considered in *Example (3)*.

Examples

(1) *The Optimal Second-order System with Noise in the Communications Channel*

Consideration is given to a control system which has been thoroughly studied for the case of no noise; the block diagrams of this are given in *Figures 1* and *2*.

The equation of the optimal switching line of the system without interference has the form:

$$x_0 = -\frac{x_1^2}{2kN} \operatorname{sign} x \tag{28}$$

It will be taken that under the influence of interference g the control signal is able at each moment of time to adopt independently one of the following values

$$v = \begin{cases} a_1 u & \text{with probability } p_1 \\ a_2 u & \text{with probability } p_2 \\ \cdots \quad \cdots & \\ a_m u & \text{with probability } p_m \end{cases} \tag{29}$$

(u adopts the value $\pm N$). In this case

$$\gamma = \sum_{i=1}^{m} a_i p_i,$$

and in accordance with (25) the equation of the optimal switching line in the system with interference (29) has the form

$$x_0 = -\frac{x_1^2}{2kN \sum_{i=1}^{m} a_i p_i} \operatorname{sign} x_1 \tag{30}$$

(2) *The Optimal Second-order Control System with a Digital Computer Inside the Control Loop*

Consideration is given to a second-order plant control system, a block diagram of which is given in *Figure 3*. The optimal switching-line equation[9] has the form:

$$\left[1 + \frac{\alpha}{F(1 - a_t/K_1 F)}\right]^{T_1} = \left[1 + \frac{\beta}{F(1 - b_t/K_2 F)}\right]^{T_2} \tag{31}$$

here a_t and b_t are the values of a and b at the end of the second section of the optimal trajectory,

$$\alpha = \frac{a_t - a}{K_1}, \quad \beta = \frac{b_t - b}{K_2}$$

F is the amplifier saturation level.

It is assumed that the optimal control of the plant (*Figure 3*) is realized in the loop containing the digital computer, so that the coordinates α and β are determined with an error, and the values $\alpha + n_\alpha$ and $\beta + n_\beta$ reach the controller input.

The random signals $n\alpha$ and n_β are, for example, quantization noises and on each interval adopt one of the evenly distributed values with a probability density:

$$p(n_{\alpha,\beta}) = \begin{cases} \dfrac{1}{2\Delta_{\alpha,\beta}} & \text{if } |n_{\alpha,\beta}| \leq \Delta_{\alpha,\beta} \\ 0 & \text{if } |n_{\alpha,\beta}| > \Delta_{\alpha,\beta} \end{cases} \tag{32}$$

The optimal switching-line equation, in accordance with (27) has the form

$$\left[1 + \frac{\alpha + \Delta_\alpha \operatorname{sign} \alpha}{F(1 - a_t/K_1 F)}\right]^{T_1} = \left[1 + \frac{\beta - \Delta_\beta \operatorname{sign} \beta}{F(1 - b_t/K_2 F)}\right]^{T_2} \tag{33}$$

(3) *The Second-order System with Noise in Channel H*

The block diagram in *Figure 4* is considered. Here in the channel serving for metering the coordinate x_i the additive interference η is a Gaussian noise with zero mean value:

$$p(\eta) = \frac{1}{\sigma_\eta \sqrt{2\pi}} \exp\left\{-\frac{\eta^2}{2\sigma_\eta^2}\right\} \tag{34}$$

It is required to ensure time optimality with accuracy of no less than 5 per cent with an aperiodic transient process.

It is known that transient processes without overshoot in the system under review correspond to the switching lines

$$x_0 = \frac{x_1^2}{akN} \operatorname{sign} x_1 \tag{35}$$

when $a \leq 2$, while a 5 per cent extension of the transient process in the processing of step signals is obtained when $a = 1{\cdot}650$.

In a relay control system the coefficient $\gamma(x)$ is determined by the equation

$$\gamma(x) = 1 - 2\,p_H(x) \tag{36}$$

where $p_H(x)$ is the probability of wrong choice of control action at the point x:

$$p_H(x) = P\{\operatorname{sign} u^0(x) = -\operatorname{sign} u(x)\}$$

The magnitude of $\gamma(x)$ rises as the distance $r_1 = x_{1n} - x_1$ increases, and at some r_1^* becomes greater than 0·825. Here x_{1n} is the coordinate x, of a point lying on the switching line and having a coordinate identical with the point under consideration $x_{0n} = x_0$:

$$\psi(x_{0n}, x_{1n}, N) = 0$$

For all the points x, for which $r_1 > r_1^*$, the magnitude of $\gamma(x)$ will be replaced by $\gamma^* = 0{\cdot}825$. The resultant control system with interference will possess the property, that for all x $\gamma(x) \leq 0{\cdot}825$, while for $r_1 < r_1^*$, $\gamma < 0{\cdot}825$.

The reduction of $\gamma(x)$ when $r_1 < r_1^*$ stems from the presence of the constrained interference η^* which only manifests itself when 1 $r_1 < r_1^*$

$$|\eta^*| < r_1^* \tag{37}$$

The examination of a system of control of a plant has been arrived at with the equation k/p^2 under the constraint

$$|u| \leq 0{\cdot}825\, N$$

[in such a system the switching-line equation has just the form of (35)], while on the system there acts the constrained interference (37). It only remains to find the magnitude of r_1^* and substitute it into eqn (27).

From the general formula

$$P\{\alpha < \eta < \beta\} = \frac{1}{2}\left[\Phi\left(\frac{\beta}{\sigma_\eta\sqrt{2}}\right) - \Phi\left(\frac{\alpha}{\sigma_\eta\sqrt{2}}\right)\right]$$

taking (36) into account, one obtains

$$\gamma = \Phi\left(\frac{r_1^*}{\sigma_\eta\sqrt{2}}\right)$$

and for $\gamma = 0{\cdot}825$ one has $r_1^* = 1{\cdot}343\,\sigma\eta$.

The equation of a switching line which is optimal with accuracy up to 5 per cent has the form:

$$x_0 = -\frac{(x_1 + 1{\cdot}343\,\sigma_\eta \operatorname{sign} x_1)^2}{1{\cdot}650\,kN} \operatorname{sign} x_1 \tag{38}$$

Results

Figures 5 and 6 show the graphs of performance of a step signal of 20 V amplitude by a system controlling a plant $1/p_2$, $N = 20$ V.

Figure 5 corresponds to *Example 1*, and in *Figures 5(a)* and *5(b)* $\gamma = 0{\cdot}645$. The switching line equation in *Figures 5(a)* and *5(b)* has the form:

$$x_0 = -\frac{x_1^2}{40} \operatorname{sign} x_1 \tag{39}$$

The optimal transient process (*Figure 5*) is ensured in a system with interference following the choice of the switching line

$$x_0 = -\frac{x_1^2}{25{\cdot}8} \operatorname{sign} x_1$$

Figures 6(a), *(b)* and *(c)* illustrate *Example 3* for $\varepsilon = 4$ per cent, $\gamma^* = 0{\cdot}90$. *Figure 6(a)* shows the performance of a step signal of amplitude 20 V without interference with switching line (39). *Figure 6(b)* demonstrates the performance of the signal for $\sigma\eta = 14{\cdot}3$ V and a switching line which is optimal without interference. The optimal (with accuracy up to 4 per cent) transient process without overshoot is shown in *Figure 6(c)*.

The switching-line equation is

$$x_0 = -\frac{(x_1 + 23{\cdot}6 \operatorname{sign} x_1)^2}{36} \operatorname{sign} x_1$$

The frequency band of the noise signal $f\eta_1$ is 10 c/sec.

References

1 KRASOVSKY, N. N. Optimal control with random disturbances. *Priklad. Mat. mekh.* 24, No. 1 (1960)

2 AOKI, M. Stochastic time-optimal systems. *Appl. Ind.*, No. 54 (May 1961)

3 FELDBAUM, A. A. Problems of the statistical theory of automatic optimization systems. *Automatic and Remote Control*, Vol. 2. 1960. London; Butterworths

4 NOVOSELTSEV, V. N. Optimal control in a second-order relay sampled-data system with random disturbances. *Automat. telemech.* 22, No. 7 (1961)

5 FLORENTIN, J. J. Optimal control of continuous time, Markov and stochastic systems. *J. Electron Contr.*, 1st Series. 10, No. 6 (1961)

6 NOVOSELTSEV, V. N. Time optimal control systems with random interference. *Automat. telemech.* 23, No. 12 (1960)

7 PONTRYAGIN, L. S. *The Mathematical Theory of Optimal Processes*, 1961. Moscow, Fizmatgiz

8 FILLIPOV, A. O. Some problems of optimal control theory. *Vestn. Moskovsk. Univ.* No. 2 (1959)

9 SMITH, J. M. *Feedback Control Systems*, Pt. H. 1958. New York; McGraw-Hill

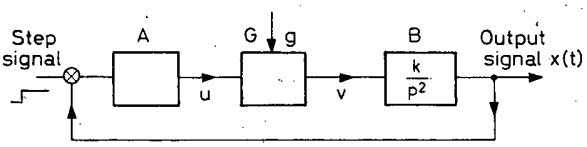


Figure 1

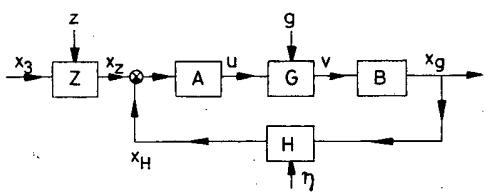


Figure 2

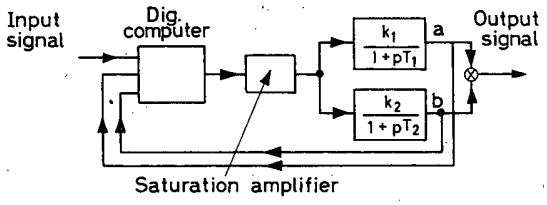


Figure 3

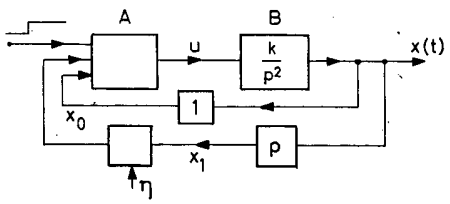
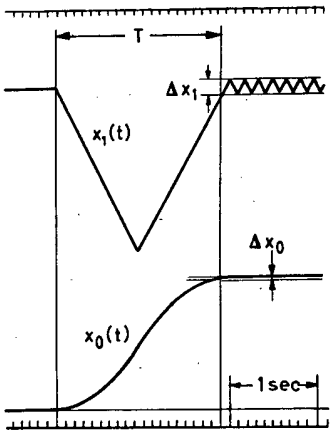
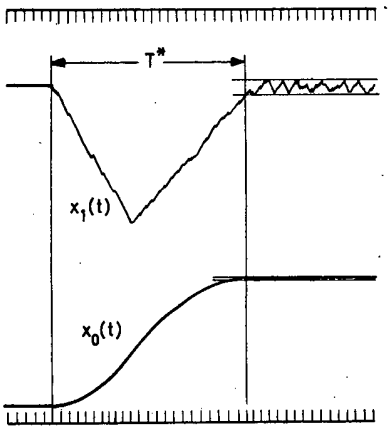


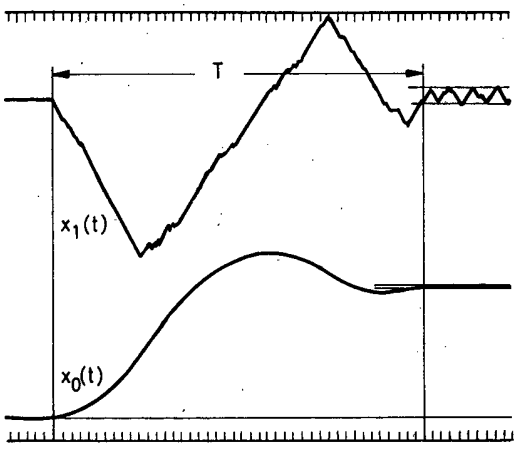
Figure 4



(a)

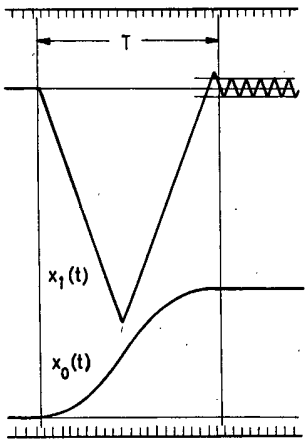


(b)

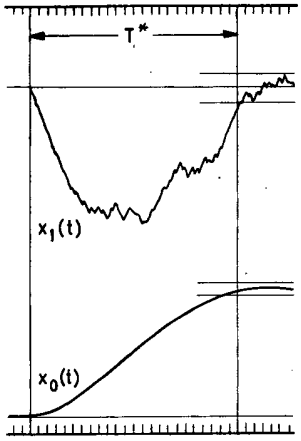


(c)

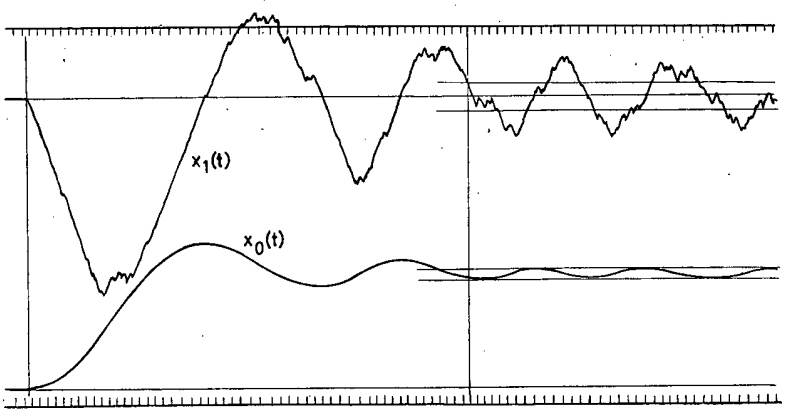
Figure 5



(a)



(b)



(c)

Figure 6

Dual Control Theory Problems

A. A. FELDBAUM

Introduction

The controlling device in an automatic system solves two problems that are closely interrelated, but which differ in character. In the first place, on the basis of information that is fed into it, it clarifies the properties and state of the controlled plant. In the second place, on the basis of the properties discovered in the plant, it determines which steps have to be taken for successful control. The first task is that of studying the plant; the second task is that of adjusting the plant to the required operating conditions. In the simplest types of systems, the solution of one of these problems may be absent or have a simple form. In complex cases, the controlling device should solve both indicated problems. Below are considered problems involved in the construction of optimal devices that solve both problems simultaneously.

Optimal systems may be divided into three types: (*a*) optimal systems having complete or the maximum information possible about the controlled plant; (*b*) optimal systems having incomplete information about the plant, and with an independent or passive storage of it in the control process; (*c*) optimal systems having incomplete information about the plant, and with an active storage of it in the control process.

In order to clarify this classification, it is necessary to determine what is meant by information regarding the controlled plant. In *Figure 1*, the controlled plant B is shown in the shape of a rectangle; in it, x is the output magnitude, u is the controlling action and z is a non-controlled disturbance. If the object has several inputs and outputs, then the magnitudes x, u, z must be considered as vectors.

The dependence

$$x = F(u, z) \tag{1}$$

may assume the form of a relationship between the values of x, u, z at the same moment of time; either the form of a differential or some other type of control. In the general case, F is an operator.

Information regarding the plant is gathered from the following elements: (*a*) information about the plant operator F; (*b*) information about the disturbance z, which acts on the plant; (*c*) information about the state of the plant—for example, about the coordinates of the point that represents the state of the plant in the phase space; and (*d*) information regarding the control purpose.

The last-mentioned information element should indicate the ideal that is to be attained, and also the 'price' of a deviation from this ideal. For this reason, the control goal may be conveniently presented in the form of a requirement for the minimization of some functional Q, which depends on the character of the x, u, z processes, and also on some externally assigned control x^* (of the given action).

Below is a statement limited by conditions of this type:

$$Q(x, x^*) = \min \tag{2}$$

Let Q be called the optimization criterion. Set the requirement, for example, that the ideal process x be identical with x^*, and that the 'cost' of the deviation from the ideal be expressed by the formula:

$$Q = c \int_0^T (x - x^*)^2 \, dt \tag{3}$$

where c and T are constants. Expression (3) is a partial example of the optimization criterion. A system called optimal in which the minimum of the criterion Q is realized, while fulfilling the additional conditions that characterize the problem—for example, that u and x belong to some admissible domains $R(u)$ and $R(x)$, respectively.

Complete information regarding some arbitrary dependence implies absolutely accurate knowledge of it. For example, complete information about some time function or other, $f(t)$, denotes that its values are known for any arbitrary values of t. It is considered below, that complete information is available regarding the operator F, and also regarding the optimization criterion Q. All that is unknown and unforeseen in the plant is attributed to the disturbance z, and what is unknown in the control goal—to the assigned control x^*.

Theories regarding optimal systems, with complete information about the plant, were developed in a number of papers[1,4]. In the theory regarding systems with an independent storage of information about the plant, consideration was given, for the main part, to closed-loop systems. Here statistical study methods were introduced[1, 5, 7].

Problems have been considered[8, 9] that pertain to the third type of optimal system theories. The theory involving systems of the third type contains characteristics that are common both to theories of the first as well as to those of the second type, and which therefore unite them to a certain extent. However, the third type of theory is also characterized by its own specific features.

The block diagram, previously studied[8, 9], is shown in *Figure 2*. A closed-loop system is considered, in which the operator of the plant B and the optimization criterion Q have been assigned. Plant B is acted on by a random disturbance z, which cannot be measured directly. The controlling action u is admitted from the controlling device A to the plant B, through the connecting channel G, where it is mixed with the random noise g. For this reason, the action v, at the input of the plant B, is not equal, generally speaking, to the magnitude u. Further, information regarding the plant's state x goes through the connecting channel H, where it is mixed with the random interference h. The output y of the connecting channel H is admitted to the input of the controlling device A. Action x^* is also admitted to the input of device A through channel H^*, with noise h^*.

In the circuit shown in *Figure 2*, processes are possible that have been considered in the theories of the first two types. A study of disturbance z, i.e., essentially, of the changing characteristics of plant B, may be carried out in the circuit of *Figure 2*, not by means of passive observation, but through an active method, by means of rational experiments. The plant is 'felt', as it were, by the u actions, which have a preceptional character, and the results y of these actions are analysed by the A device. The purpose of these actions is to promote a more rapid and accurate study of the plant's characteristics; this can develop a better principle for controlling it.

However, the controlling action is necessary not only for the purpose of study, but also for directing and adjusting the plant to the required operating conditions. Consequently, in the circuit in *Figure 7*, the controlling actions should have a dual character; to a certain degree they should be studying actions, but to another degree also directing actions. That is why the theory underlying systems of this type is called a dual-control theory.

It is precisely this duality of control that constitutes the physical fact distinguishing the third type of optimal systems from the first two. In the first one, dual control is not necessary, in so far as the controlling device without it possesses complete (or maximally possible) information about the plant. In the second type of system, dual control is impossible, because the information is stored by means of observation alone, and the rate of its storage does not depend at all on the strategy of the controlling device.

Setting Up the Problem

Theoretical problems are considered below only for systems that are time discrete. All the magnitudes indicated in the circuit in *Figure 2* are considered only for discrete time moments, $t=0\cdot1, 2, \ldots, n$, where n has been fixed. The value of any arbitrary magnitude for an S discrete time moment has been provided with an index S—for example, x_s^*, x_s, y_s, etc. The transmission lines of all magnitudes are assumed to be single-channelled, and the plant B is considered to have no memory. Therefore, its equation may be written thus:

$$x_s = F(v_s, z_s) \tag{4}$$

A generalization of the conclusion set forth below, for more complex plant cases, with several inputs and outputs, and also for plant having a memory, may be carried out in the same way as was done previously[8].

Let h_s^*, h_s and g_s represent a series of independent, random magnitudes with invariable distribution densities: $P(h_s^*)$, $P(h_s)$, $P(g_s)$. Further, let:

$$z_s = z_s(s, \bar{\mu}_s) \tag{5}$$

and

$$x_s^* = x_s^*(s, \bar{\lambda}_s) \tag{6}$$

where the average μ_s and average λ_s, in contrast with refs. 8 and 9, are not random magnitudes, but discrete vector Markov random processes. In other words, the average μ_s and average λ_s are vectors:

$$\bar{\mu}_s = (\mu_s^1, \ldots, \mu_s^m) \tag{7}$$

and

$$\bar{\lambda}_s = (\lambda_s^1, \ldots, \lambda_s^l) \tag{8}$$

μ_s^i and μ_s^j, in the general case, are mutually interrelated, scalar Markov discrete processes. The same holds for λ_s^i and λ_s^j. However, the vectors average μ_s, average λ_s, and also, noises h_s^*, h_s, g_s are considered independent.

Consider the Markov process characteristics average μ_s and average λ_s as having been given. This implies that one has been given, as the initial probability densities, P_0 (average μ_0) and P_0 (average λ_0) where $t=0$, as well as the transient probability densities P (average μ_{i+1}/average μ_i) and P (average λ_{i+2}/average λ_i).

The methods for combining the signal and the noise in blocks H^*, H and G are considered as known and invariable, and the blocks themselves as having no memory. Therefore:

$$v_s = v_s(u_s, g_s);\ y_s^* = y_s^*(h_s^*, x_s^*);\ y_s = y_s(h_s, x_s) \tag{9}$$

The control goal is determined in the following manner: let the specific loss function (the 'cost' of deviation from the ideal), which corresponds to the s time moment, have the form:

$$W_s = W_s(s, x_s, x_s^*) \tag{10}$$

Further, let the overall loss function W, for the entire time period n, be equal to the sum of the specific loss functions:

$$W = \sum_{s=0}^{s=n} W_s(s, x_s, x_s^*) \tag{11}$$

A system is called optimal for which the average risk R (mathematical expectation M of the magnitude W) is minimal. The risk magnitude is expressed by the formula:

$$R = M\{W\} = M\left\{\sum_{s=0}^{s=n} W_s(s, x_s, x_s^*)\right\} = \sum_{s=0}^{s=n} M\{W_s\} = \sum_{s=0}^{s=n} R_s \tag{12}$$

The expression $R_s = M(W_s)$ is called the specific risk in the s cycle. The magnitude R plays the part, here, of the optimization criterion Q.

Introduce the tentative vectors ($0 \leq s \leq n$):

$$\left.\begin{aligned} \vec{u}_s &= (u_0, u_1, \ldots, u_s); & \vec{x}_s^* &= (x_0^*, x_1^*, \ldots, x_s^*) \\ \vec{v}_s &= (v_0, v_1, \ldots, v_s); & \vec{y}_s &= (y_0, y_1, \ldots, y_s) \\ \vec{x}_s &= (x_0, x_1, \ldots, x_s); & \vec{y}_s^* &= (y_0^*, y_1^*, \ldots, y_s^*) \end{aligned}\right\} \tag{13}$$

and the matrices of average vector μ_s, average vector λ_s, which are made up of the vector columns of average μ_s, average λ_s:

$$\bar{\bar{\mu}}_s = (\bar{\mu}_0, \bar{\mu}_1, \ldots, \bar{\mu}_s);\quad \bar{\bar{\lambda}}_s = (\bar{\lambda}_0, \bar{\lambda}_1, \ldots, \bar{\lambda}_s) \tag{14}$$

Consider that the control device, in the general case, possesses a memory. In addition to this, assume, for general purposes, that the algorithm of this device is a random one. The term 'random strategy' is also employed. This implies that the value u_s is a random function of the magnitudes y_i and [redacted] which were admitted to the input of device A during the preceding moments of time, of u_i ($i < s$), and also of the values y_j^* ($j \leq s$). It is required to find the optimal probability densities:

$$P_s(u_s) = \Gamma_s(u_s | \vec{u}_{s-1}, \vec{y}_{s-1}, \vec{y}_s)_{(0 \leq s \leq n)} \tag{15}$$

The problem consists in finding such a series of functions Γ_s, in the case of which the average risk R will be minimal. Inasmuch as Γ_s is the probability density, therefore:

$$\int_{R(u_s)} \Gamma_s(u_s)\,d\mathscr{R}(u_s)=1 \qquad (16)$$

where $\mathscr{R}(u_s)$ designates the region for the magnitude changes u_s, and $d\mathscr{R}(u_s)$ represents its infinitely small element. And thus it must be found that the optimal functions $\Gamma_s > 0$, which are limited by condition (16).

Derivation of the Basic Formula

First write the formula for a conditional, specific risk, r_s, understanding by this a risk in the s cycle, with a fixed 'prehistory' of the control device inputs, i.e., with fixed values for $\vec{y}_s^*, \vec{y}_{s-1}, \vec{u}_{s-1}$:

$$\tau_s = M\{W_s|\vec{y}_s^*, \vec{u}_{s-1}, \vec{y}_{s-1}\}$$
$$= \int_{R(\lambda_s, x_s)} W_s[s, x_s, x_s^*(s_1, \bar{\lambda}_s)]\cdot P(\bar{\lambda}_s, x_s|\vec{y}_s^*, \vec{u}_{s-1}, \vec{y}_{s-1})\,dR(\bar{\lambda}_s x_s) \qquad (17)$$

Here $\mathscr{R}$ (average, λ_s, x_s) is the domain of changes for average λ_s and x_s, and $d\mathscr{R}$ (average λ_s, x_s) is its infinitely small element; P (average λ_s x_s/vector y_s^*, vector u_{s-1}, vector y_{s-1}) is the conditional, common probability density of the average λ_s and x_s, with fixed vectors y_s^*, u_{s-1}, y_{s-1}. In conformity with a well-known theorem of the theory of probabilities, an equality exists:

$$P(\bar{\lambda}_s, x_s|\vec{y}_s^*, \vec{u}_{s-1}, \vec{y}_{s-1})$$
$$= P(\bar{\lambda}_s|\vec{y}_s^*, \vec{u}_{s-1}, \vec{y}_{s-1})\cdot P(x_s|\vec{y}_s^*, \vec{u}_{s-1}, \vec{y}_{s-1}, \vec{\lambda}_s)$$
$$= P(\bar{\lambda}_s|\vec{y}_s^*)\cdot P(x_s|\vec{y}_s^*, \vec{u}_{s-1}, \vec{y}_s) \qquad (18)$$

The last transposition is accurate, because the probability density of average λ_s, with a fixed vector y_s^*, will not change if vector u_{s-1}, vector y_{s-1} are also fixed (see *Figure 4*). Further, the probability density of x_s with a fixed vector y_s^*, will not change, if in addition average λ_s is fixed. The second multiple (18) is rewritten in an expanded form:

$$P(x_s|\vec{y}_s^*, \vec{u}_{s-1}, \vec{y}_{s-1})$$
$$= \int_{R(\bar{\mu}_s, \bar{u}_s)} P(x_s|\bar{\mu}_s, u_s)\,P_s(\bar{\mu}_s)\,\Gamma_s(u_s|\vec{y}_s^*, \vec{u}_{s-1}, \vec{y}_{s-1})\,d\mathscr{R}(\bar{\mu}_s, \bar{u}_s) \qquad (19)$$

where $\mathscr{R}$ (average μ_s, average u_s) is the domain of changes for average μ_s and average u_s, and P_s (average μ_s) is the *a posteriori* probability density of average μ_s in the s cycle:

$$P_s(\bar{\mu}_s) = P(\bar{\mu}_s|\vec{u}_{s-1}, \vec{y}_{s-1})$$
$$= \int_{R(\bar{\mu}_{s-1})} P(\bar{\mu}_s|\bar{\mu}_{s-1})\,P(\bar{\mu}_{s-1}|\vec{y}_{1-s}, \vec{u}_{s-1})\,d\mathscr{R}(\bar{\mu}_{s-1}) \qquad (20)$$

Inasmuch as:

$$P(\bar{\mu}_{s-1}|\vec{y}_{s-1}, \vec{u}_{s-1})$$
$$= \int_{R(\vec{\mu}_{s-2})} P(\vec{\mu}_{s-1}|\vec{y}_{s-1}, \vec{u}_{s-1})\,d\mathscr{R}(\vec{u}_{s-2}) \qquad (21)$$

where $\mathscr{R}$ (average vector μ_{s-2}) is the domain of changes for the matrix of average vectors (μ_{s-2}), therefore, it is necessary to find the conditional probability density of P (average vector μ_{s-1}, vector y_{s-1}), for the matrix of average vectors μ_{s-1}. From the equality:

$$P(\vec{\bar{\mu}}_{s-1}, \vec{u}_{s-1}, \vec{y}_{s-1})$$
$$= P(\vec{u}_{s-1}, \vec{y}_{s-1}|\vec{\bar{\mu}}_{s-1})\cdot P(\vec{\bar{\mu}}_{s-1})$$
$$= P(\vec{\bar{\mu}}_{s-1}|\vec{u}_{s-1}, \vec{y}_{s-1})\cdot P(\vec{u}_{s-1}, \vec{y}_{s-1}) \qquad (22)$$

is found:

$$P_{s-1}(\vec{\bar{\mu}}_{s-1}) = P(\vec{\bar{\mu}}_{s-1}|\vec{u}_{s-1}, \vec{y}_{s-1})$$
$$= \frac{P(\vec{u}_{s+1}, \vec{y}_{s-1}|\vec{\bar{\mu}}_{s-1})\,P(\vec{\bar{\mu}}_{s-1})}{P(\vec{u}_{s-1}, \vec{y}_{s-1})} \qquad (23)$$

Here P (vector u_{s-1}, vector y_{s-1}) is the common *a priori* probability density of vectors u_{s-1}, y_{s-1}; P (average vector μ_{s-1}) is the *a priori* probability density of the matrix of average vectors μ_{s-1}, and P (vector u_{s-1}, vector y_{s-1}/average vector μ_{s-1}) is the conditional probability density of vectors u_{s-1}, y_{s-1}, with a fixed matrix of average vectors μ_{s-1} (the probability function). In passing $(s-1)$ times around the closed logs in *Figure 4*, it is possible, as in the derivation[8, 9] to find the expression:

$$P_{s-1}(\vec{\bar{\mu}}_{s-1}) = \frac{P_0(\mu_0)\prod_{i=1}^{s-1} P(\bar{\mu}_i|\bar{\mu}_{i-1})\left[\prod_{i=0}^{s-1} P(y_i|\mu_i, i, u_i)\right]\left[\prod_{i=0}^{s-1}\Gamma_i\right]}{P(\vec{u}_{s-1}, \vec{y}_{s-1})} \qquad (24)$$

The substitution of (24) in (21), and further on, of (21) in (20), (20) in (19) and (19) in (18) will make it possible to determine the second co-factor in (18). Now consider the first co-factor of this expression—the *a posteriori* probability density $\bar{\lambda}_s$:

$$P_s(\bar{\lambda}_s) = P(\bar{\lambda}_s|\vec{y}_s^*)$$
$$= \int_{R\vec{\lambda}(s-1)} P(\bar{\bar{\lambda}}_s|\vec{y}_s^*)\,d\mathscr{R}(\vec{\bar{\lambda}}_{s-1}) \qquad (25)$$

Inasmuch as:

$$P(\bar{\bar{\lambda}}_s, \vec{y}_s^*) = P(\bar{\bar{\lambda}}_s)\cdot P(\vec{y}_s^*|\bar{\bar{\lambda}}_s) = P(\bar{\bar{\lambda}}_s|\vec{y}_s^*)\cdot P(\vec{y}_s^*)$$

therefore:

$$P_s(\bar{\bar{\lambda}}_s) = P(\bar{\bar{\lambda}}_s|\vec{y}_s^*) = P(\bar{\bar{\lambda}}_s)\cdot\frac{P(\vec{y}_s^*|\bar{\bar{\lambda}}_s)}{P(\vec{y}_s^*)} \qquad (26)$$

The *a priori* probability density of the matrix for average vectors λ_s is determined from the formula that is accurate for the Markov process:

$$P(\bar{\bar{\lambda}}_s) = P(\bar{\lambda}_0, \bar{\lambda}_1, \ldots, \bar{\lambda}_s)$$
$$= P_0(\bar{\lambda}_0)\cdot P(\bar{\lambda}_1|\bar{\lambda}_0)\cdot P(\bar{\lambda}_2|\bar{\lambda}_1)\ldots P(\bar{\lambda}_s|\bar{\lambda}_{s-1})$$
$$= P_0(\bar{\lambda}_0)\prod_{i=1}^{s} P(\bar{\lambda}_i|\bar{\lambda}_{i-1}) \qquad (27)$$

Further, the conditional probability density is:

$$P(\vec{y}_s^*|\bar{\bar{\lambda}}_s) = P(y_0^*|\lambda_0)\cdot P(\lambda_1^*\bar{\lambda})\ldots P(y_s^*|\bar{\lambda}_s)$$
$$= \prod_{i=0}^{s} P(y_i^*|\lambda_i) \qquad (28)$$

Consequently, from (26), (27) and (28), is obtained:

$$P_s(\vec{\lambda}_s)=\frac{P_0(\vec{\lambda}_0)\prod_{i=1}^{s}P(\vec{\lambda}_i|\vec{\lambda}_{i-1})\cdot\prod_{i=0}^{s}P(y_i^*|\vec{\lambda}_i)}{P(\vec{y}_s^*)} \tag{29}$$

By substituting (29) in (25), the final formula for P_s (average λ_s) is found.

Attention is now turned to the principal difference between formula (24) and (29) for the *a posteriori* probability densities of the average vectors μ_{s-1} and λ_s. The storage of information regarding the disturbance z or the vector average μ_s, i.e., essentially, regarding the unexpected manner involving the changing characteristics of the plant, is expressed in the fact the *a priori* probability density P_0 (average λ_0) is replaced in each new cycle by the *a posteriori* densities P_s (average μ_s) [see (20), associated with the expression (23)]. From (23) and (20) it is evident that the function P_s (average μ_s), and, consequently, also the rate of information storage depends on all the preceding strategies Γ_i ($i<s$). In other words, the rate involved in studying the plant depends on how efficiently the experiments were set up with respect to studying this plant, feeding it with the u_i actions and making analysis of the plant y_i reactions to these actions. By the way in formula (23) for P_s (average vector μ_s), which is associated with (29), a dependence on the part of the information storage rate, with regard to the vector average λ_s, as against the strategies Γ_i, does not exist, i.e., the information storage process is a passive or independent one.

By carrying out all the substitutions indicated above and then substituting (18) in (17), a final formula may be arrived at for the conditional, specific risk r_s. If the values of r_s, are considered in different experiments carried out with this system, then the vectors y^*_s, u_{s-1} and y_{s-1}, which, generally speaking, are not known beforehand, may assume different values. Let P (vector y_s*, vector u_{s-1}, vector y_{s-1}) be the density of the common distribution of these vectors; in such a case:

$$P(\vec{y}_s^*,\vec{u}_{s-1},\vec{y}_{s-1})=P(\vec{u}_{s-1},\vec{y}_{s-1})\cdot P(\vec{y}_s^*|y_{s-1},\vec{u}_{s-1}) \tag{30}$$
$$=P(\vec{u}_{s-1},\vec{y}_{s-1})\,P(\vec{y}_s^*)$$

The last transposition is accurate, inasmuch as vector y_s* does not depend on vector y_{s-1} and vector u_{s-1}. In that case, the specific risk r_s, which represents the average value of r_s, where experiments have been conducted on a large scale, is determined by the formula:

$$\mathscr{R}_r=M\{r_s\}=\int_{R(\vec{y}_s^*,\vec{u}_{s-1},\vec{y}_{s-1})} r_s P(\vec{y}_s^*,\vec{u}_{s-1},\vec{y}_{s-1})\,\mathrm{d}\mathscr{R}(\vec{y}_s^*,\vec{u}_{s-1},\vec{y}_{s-1}) \tag{31}$$
$$=\int_{R(\vec{y}_s^*,\vec{u}_{s-1},\vec{y}_{s-1})} r_s P(\vec{u}_{s-1},\vec{y}_{s-1})\,P(\vec{y}_s^*)\,\mathrm{d}\mathscr{R}(\vec{y}_s^*,\vec{u}_{s-1},\vec{y}_{s-1})$$

Having substituted here the expression for r_s, the following formula is arrived at:

$$\mathscr{R}_r=\int_{R(\vec{\lambda}_s,\vec{\mu}_s,\vec{x}_s,\vec{y}_s^*,\vec{u}_s,\vec{y}_{s-1})} W_s[s,x_s^*(s,\vec{\lambda}_s)\,x_s]\cdot P_0(\vec{\lambda}_0)\cdot\prod_{i=1}^{s}P(\vec{\lambda}_i|\vec{\lambda}_{i-1})\cdot \tag{32}$$
$$\cdot\prod_{i=0}^{s}P(y_i^*|\lambda_i)\cdot P_0(\vec{\mu}_0)\cdot\prod_{i=1}^{s}P(\vec{\mu}_i|\vec{\mu}_{i-1})\cdot\prod_{i=0}^{s-1}P(y_i|\vec{\mu}_i,i,u_i)$$
$$\cdot\prod_{i=0}^{s}\Gamma_i\,\mathrm{d}\mathscr{R}(\vec{\lambda}_s,\vec{\mu},x_s,\vec{y}_s^*,\vec{u}_s,\vec{y}_{s-1})$$

It is important to note that, although in the given case, object B has no memory, nevertheless risk R_s, in an s cycle, depends on all the Γ_i strategies at the time moments $t=0,1,\ldots,s$. The physical reason for this phenomenon, which is absent in a closed loop system, is found precisely in the duality of control. Control at a k moment of time should be calculated not only with a view towards decreasing the specific risk R_k, which corresponds to this moment of time, but also towards promoting a risk reduction, R_i ($i>k$), during the following moments of time, by means of a better study of the plant.

Determination of the Optimum Strategy

In determining the optimum strategy[8, 9] our thoughts are drawn towards dynamic programming[1]. Therefore introduce some auxiliary functions α_k ($0\le k\le n$):

$$\alpha=\alpha_k(\vec{y}_k^*,u_k,\vec{u}_{k-1},\vec{y}_{k-1})$$
$$=\int_{R(\vec{\lambda}_k,\vec{\mu}_k,x_k)} W_k[k,x_k^*(k,\vec{\lambda}_k)\,x_k]\cdot P_0(\vec{\lambda}_0)\cdot\prod_{i=1}^{k}P(\vec{\lambda}_i|\vec{\lambda}_{i-1})$$
$$\cdot\prod_{i=0}^{k}P(y_i^*|\vec{\lambda}_i)\cdot P_0(\vec{\mu}_0)\cdot\prod_{i=1}^{k}P(\vec{\mu}_i|\vec{\mu}_{i-1}) \tag{33}$$

Also, let:

$$\prod_{i=0}^{k-1}P(y_i|\vec{\mu}_i,i,u_i)\,\mathrm{d}\mathscr{R}(\vec{\lambda}_k,\vec{\mu}_k,x_k) \qquad (k=0,1,\ldots,r) \tag{34}$$
$$\beta_k=\prod_{i=0}^{k}\Gamma_i$$

In that case, the formula for the risk R_n, which corresponds to the moment of time $t=n$, will assume the form:

$$R_n=\int_{R(\vec{y}_n^*,\vec{u}_n,\vec{y}_{n-1})}\alpha_n(\vec{y}_r^*,u_r,\vec{u}_{n-1},\vec{y}_{n-1})\,\beta_{n-1}\cdot\Gamma_r\,\mathrm{d}R(\vec{y}_n^*,\vec{u}_n,\vec{y}_{n-1}) \tag{35}$$
$$=\int_{R(\vec{y}n^*,\vec{u}_{n-1},\vec{y}_{n-1})}\beta_{n-1}\,\chi_r(\vec{y}_r^*,\vec{u}_{n-1},\vec{y}_{n-1})\,\mathrm{d}\mathscr{R}(\vec{y}_n^*,\vec{u}_{n-1},\vec{y}_{n-1})$$

where:

$$\chi_r(\vec{y}_n^*,\vec{u}_{n-1},y_{n-1})$$
$$=\int_{R(u_n)}\alpha_r(\vec{y}_n^*,u_r,\vec{u}_{n-1},\vec{y}_{n-1})\,\Gamma_n(\vec{y}_n^*,u_n,\vec{u}_{n-1},\vec{y}_{n-1})\,\mathrm{d}\mathscr{R}(u_n) \tag{36}$$

On the basis of the theorem regarding average value, taking (16) into consideration the following can be written:

$$\chi_n=(\alpha_n)_\varphi\int_{R(u_n)}\Gamma_n\,\mathrm{d}\mathscr{R}(u_n)=(\alpha_n)_\varphi\ge(\alpha_n)_{\min} \tag{37}$$

Assume that all Γ_i ($i<n$) are given and that the control process, right down to the moment $t=n$, has been realized. The selection of Γ_n must be in such a way as to minimize R_n. This may be accomplished if, for any arbitrary vectors y_n*, u_{n-1}, y_{n-1}, Γ_n is selected in such a manner as to have function χ_n minimal. Let γ_n equal α_n and u_n* be the value u_n that minimizes α_n.

(38)

Evidently, u_n^* is the function of vectors $\vec{y}_n^*$, $\vec{u}_{n-1}$ and $\vec{y}_{n-1}$:

$$u_r^* = u_r^*(\vec{y}_n^*, \vec{u}_{n-1}, \vec{y}_{n-1}) \tag{39}$$

In that case, the optimum strategy Γ_r^* is given by the expression:

$$\Gamma_r^* = \delta(u_n - u_n^*) \tag{40}$$

where δ is the unit impulse function. This denotes that Γ_n is the regular strategy, and not the random one, in which case the optimal value $u_n = u_n^*$. From (39) it is evident tha the optimal value depends on the values previously observed by the control device A: u_s, y_s ($s = 0, 111, n-1$), and also on: y_i^* ($i = 0, \ldots, n$).

It is very simple to prove the accuracy of expression (40). By substituting it in formula (35), one obtains, by virtue of the known property of the δ function:

$$\chi_r = \min_{u_n \varepsilon \Omega(u_n)} \alpha_n(u_n, \vec{y}_r^*, \vec{u}_{n-1}, \vec{y}_{n-1}) = (\alpha_n)_{\min} \tag{41}$$

But, according to (34), this is actually the lowest possible value for χ_n. Consequently, Γ_n^* represents the optimum strategy.

In order to find the optimum strategies Γ_n^*, where $i < n$, one must shift gradually from the terminal moment $t = n$ to the beginning—see references 8 and 9.

As a result, the rule stated below is arrived at for the determination of the optimum strategy Γ_{n-k}^*. Introduce the function:

$$\gamma_{n-k} = \gamma_{n-k}(\vec{y}_{n-k}^*, \vec{u}_{n-k-1}, \vec{y}_{n-k-1}) = \alpha_{n-k} + \int_{R(y_{n-k}, y_{n^*-k+1})} \gamma_{n-k+1}^*(u_{n-k+1}^*, \vec{y}_{n-k+1}, \vec{u}_{n-k}, \vec{y}_{n-k})\, d\Omega(y_{n-k}, y_{n-k+1}^*) \tag{42}$$

The magnitude $\gamma_n^* = \alpha_n^*$, according to (38). Now find the value that minimizes the function γ_{n-k}, in which case:

$$\gamma_{n-k}^* = \min_{u_{n-k} \varepsilon R(u_{n-k})} \gamma_{n-k} = \gamma_{n-k}(u_{n-k}^*) \tag{43}$$

Evidently,

$$u_{n-k}^* = u_{n-k}^*(\vec{y}_{n-k}^*, \vec{u}_{n-k-1}, \vec{y}_{n-k-1}) \tag{44}$$

In that case, the optimum strategy is:

$$\Gamma_{n-k}^* = \delta(u_{n-k} - u_{n-k}^*) \tag{45}$$

i.e., the optimum strategy is regular and consists of the selection: $u_{n-k} = u_n^{*}{}_{-k}$. From (44) it is evident that $u_n^{*}{}_{-k}$ depends on the values of u_i and y_i, which had been observed by the control device during the preceding moments, where $i < n - k$, and also y_j^* ($j \leq n - k$). Consequently, algorithm (44) is realized physically.

In the partial case, when the average λ_s process is converted to a random value of average λ, and average μ_s to a random value of average μ, formula (33) for α_k is simplified and assumes the form:

$$\alpha_k = \int_{R(\bar{\lambda}, \bar{\mu}, x_k)} W_k[k, x_k^*(k, \bar{\lambda}), x_k] P_0(\bar{\lambda}) \prod_{i=0}^{k} P(y_i^* | \bar{\lambda}) \cdot P_0(\bar{\mu}) \prod_{i=0}^{k-1} P(y_i | \bar{\mu}, i, u_i)\, d\mathscr{R}(\bar{\lambda}, \bar{\mu}, x_k) \tag{46}$$

If x_k^* are given in advance, then the formula proves to be still simpler:

$$\alpha_k = \int_{R(\bar{\mu}, x_k)} W_k(k, x_k) P_0(\bar{\mu}) \prod_{i=0}^{k-1} P(y_i | \bar{\mu}, i, u_i)\, dR(\bar{\mu}, x_k) \tag{47}$$

These formulae had been previously brought out[8, 9].

Examples

Consider three examples that illustrate the above theory. *Figure 5*, shows a representation of the simplest system for which $h_s = 0$, μ is the random magnitude and the equation has the form:

$$\left.\begin{aligned} v_s &= u_s + y_s \\ y_s^* &= \alpha_s^* + h_s^* \\ x_s &= v_s + \mu = u_s + g_s + \mu \end{aligned}\right\} \tag{48}$$

Let the r and om magnitudes μ, g_s and h_s^* have normal distribution rules, with 0 average values and σ_μ^2, σ_g^2 and σ_h^2 dispersions, respectively. Further, λ_0 and σ_λ are known magnitudes, in which case,

$$x_s^* = \lambda = \text{const}; \quad P(\lambda) = \frac{1}{\sigma_\lambda \sqrt{2\pi}} \exp\left\{ -\frac{(\lambda - \lambda_0)^2}{2\sigma_\lambda^2} \right\} \tag{49}$$

Let:

$$W_s = W_s(s, x_s, x_s^*) = (x_s - x_s^*) = (x_s - \lambda)^2 \tag{50}$$

As a result of basing the solution on the method described above (see reference 9) and with the application of formula (46), the optimal control rule in the following form is found:

$$u_s^* = \frac{\lambda_0}{1 + \left(\frac{\sigma_\lambda}{\sigma_\lambda}\right)^2 (s+1)} + \frac{\sum_{i=0}^{s} y_i^*}{\left(\frac{\sigma_h}{\sigma_\lambda}\right)^2 + (s+1)} - \frac{\sum_{i=0}^{s-1} (x_i - u_i)}{s + \left(\frac{\sigma_y}{\sigma_\mu}\right)^2} \tag{51}$$

The explain the meaning of this formula, if interferences g_s and h_s^* were absent, then, for the purpose of obtaining an ideal value, $x_s = x_s^* = \lambda$, that would assure the magnitude $W_s = 0$, it would be necessary to establish the value of $u_s = u_s^* - \mu = \lambda - \mu$. In formula (51), the first two components yield an evaluation for λ on the basis of the observed y_i^* values. The final term yields an evaluation for μ on the basis of the observed differences $(x_i - u_i)$. It is evident from *Figure 5*, in fact, that $x_i - u_i = \mu + y_i$. Consequently, a neutralization of the differences $(x_i - u_i)$ yields an evaluation for the magnitude μ. With sufficiently high values for s, the final term of expression (51) is approximately equal to the arithmetic mean of the values $(x_i - u_i)$.

Consider another example pertaining to the same circuit in *Figure 5*. Let $h_s^* = 0$, while μ is replaced by μ_s and represents a gaussian, discrete, Markov random process; in such case,

$$P_0(\mu_0) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp\left\{ -\frac{\mu_0^2}{2\sigma_0^2} \right\}$$
$$P(\mu_k | \mu_{k-1}) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp\left\{ -\frac{(\mu_k - \mu_{k-1})^2}{2\sigma_1^2} \right\} \tag{52}$$

The magnitudes g_s and W_s are the same as in the preceding example, and $x_s^* = x^*$ is a known constant. In that case (see reference 10) the optimal control rule has the form:

$$u_k^* = x^* - \sum_{i=0}^{k-1} e_{i,k}(x_i - u_i) \tag{53}$$

The second component of formula (53) represents an evaluation for the magnitude μ_s. The values of the weighting

coefficients, $l_{i,k}$, which are computed from comparatively complex formulae that are not set forth here, possess the property:

$$\frac{l_{i\,k}}{l_{j,\,k}}<1 \quad (0\leq i<j\leq k) \tag{54}$$

The physical significance of this property in the optimum strategy (53) consists in that a lesser weight is imparted to information of older origin, inasmuch as it 'becomes obsolete'. Thus, there takes place, in control device A, not only a process of storing new information, but also a process of degrading obsolete information.

For the established process, where k tends to infinity in formula (53), the $l_{i,k}$ coefficients diminish in accordance with the law of geometric progression as the value $v = k - i$ increases, i.e., as the previously measured difference $(x_i - u_i)$ is withdrawn from the current moment of time. It is not difficult to realize such an algorithm by means of the simplest circuit.

The examples given above are degenerate, since they are equivalent to examples in which the value of the unknown parameter μ is measured with a certain amount of error. The above theory, however, by means of a uniform method, makes it possible to examine even more complicated problems. Consider the system that is represented in *Figure 6*. The equations for this system have the form:

$$x_s=h_s^2=(u_s+\mu)^2, \quad y_s=x_s+h_s \tag{55}$$

The magnitudes g_s and $h_s^* = 0$. Noise h_s has a normal distribution with a zero average value and a σ_n^2 dispersion. The magnitude x_s^* is absent (for example, it may be assumed that $x_s^* = 0$). All the values of $W_s = 0$, with the exception of the last one:

$$W_r=x_r \tag{56}$$

Assume that μ is a random magnitude with a probability density of $P_0(\mu)$, in which case $|\mu| \leq 1$. In the same manner, it may also he assumed that $|u| \leq 1$. The problem consists in determining the optimum algorithm for the control device A that would satisfy the condition:

$$R=M\{W_r\}=M\{x_r\}=\min \tag{57}$$

The solution of this problem produces the optimal method for finding the minimum of the parabolic function $x = (u + \mu)^2$, where μ is unknown, and x is measured with an h error. At first, where $i = 0, 1, \ldots, n-1$, tentative values for u_1 are established, and the corresponding magnitudes y_i are measured. Following this, where $i = n$, such a u_n is established as to give a minimal mathematical anticipation to the x_n value that corresponds to it.

For the given problem:

$$P(y_i|\mu,i,u_i)=\frac{1}{\sigma_r\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma_r^2}[y_i-u_i^2-2u_i\mu-\mu^2]^2\right\}$$

$$=\frac{1}{\sigma_n\sqrt{2\pi}}\exp\{a_i+b_i\mu+c_i\mu^2+d_i\mu^3+\mu^4\} \tag{58}$$

where:

$$\left.\begin{aligned} a_i&=-\frac{1}{2\sigma_r^2}(y_i-u_i^2)^2; \quad & b_i&=-2u_i(u_i^2-y_i)\frac{1}{\sigma_r^2}\\ c_i&=-\frac{1}{\sigma_r^2}(3u_i^2-y_i); \quad & d_i&=-\frac{2u_i}{\sigma_r^2}\end{aligned}\right\} \tag{59}$$

All $\alpha_k = 0$, with the exception of α_n, for which, in conformity with (47) it is found that (integration for x_n is replaced by the substitution $x_n = (u_n + \mu)^2$):

$$\alpha_r=\int_{-1}^{1}(u_r+\mu)^2P_0(\mu)\frac{1}{(\sigma_r\sqrt{2\pi})^r}$$

$$\exp\left\{\sum_{i=0}^{n-1}(a_i+b_i\mu+c_i\mu^2+d_i\mu^3+\mu^4)\right\}d\mu \tag{60}$$

By making use of this expression, it is possible to find γ_i and the values for u_i^*, which, in minimizing γ_i, prove to be optimal. This is accomplished by means of a succession of alternating minimizations and integrations, in the course of which, it is necessary to memorize the functions of three variables that are called, as is known (see, for example reference 11) sufficient coordinates, but the functions of three variables are too complex. For this reason, in the given case, the sufficient coordinates prove to be, figuratively speaking, insufficient for a convenient solution. However, the solution may be considerably simplified. As was indicated by calculations, by means of expansion into a Pike and Silverberg series[12], it is possible to assume, with a sufficient degree of accuracy:

$$b_i\mu+c_i\mu^2+d_i\mu^3+\mu^4=$$

$$\cong\varphi_1(y_i,u_i)f_1(\mu)+\varphi_2(y_i,u_i)f_2(\mu) \tag{61}$$

where $\varphi_1, f_1, \varphi_2, f_2$ are some functions. In that case,

$$\alpha_r=\int_{-1}^{1}(u_n+\mu)^2P_0(\mu)\exp$$

$$\left\{A_{n-1}+\sum_{i=0}^{n-1}[\varphi_1(y_i,u_i)f_1(\mu)+\varphi_2|y_i,u_i)f_2(\mu)]\right\}d\mu$$

$$=\exp\{A_{n-1}\}\int_{-1}^{1}(u_n+\mu)^2P_0(\mu)$$

$$\exp\{E_{n-1}f_1(\mu)+F_{n-1}f_2(\mu)\}\,d\mu \tag{62}$$

where

$$\left.\begin{aligned} A_s&=\sum_{i=0}^{s}a_i=A_{s-1}+a_s\\ E_s&=\sum_{i=0}^{s}\varphi_1(y_i,u_i)=E_{s-1}+\varphi_1(y_s,u_s)\\ F_s&=\sum_{i=0}^{s}\varphi_2(y_i,u_i)=F_{s-1}+\varphi_2(y_s,u_s)\end{aligned}\right\} \tag{63}$$

For this reason,

$$\gamma_n^*=\alpha_n^*=\min_{u_n\in\Omega(u_n)}\alpha_n=\exp\{A_{n-1}\}\,\Theta_n^*(E_{n-1},F_{n-1}) \tag{64}$$

where Θ_n^* is the function of two variables.

Further,

$$\gamma_{n-1}=\int_{-\infty}^{\infty}\gamma_n^*\,dy_{n-1}=\exp\{A_{n-2}\}\cdot\Theta_n(E_{n-2},F_{n-2},u_{n-1}) \tag{65}$$

and, if one assumes that:

$$\Theta_{n-1}^*=\min_{u_{n-1}\in\Omega(u_{n-1})}\Theta_{n-1} \tag{66}$$

$$\gamma^*_{n-1} = \exp\{A_{n-2}\} \cdot \Theta^*_{n-1}(E_{n-2}, F_{n-2}) \quad (k = 0, 1, \ldots, n) \tag{67}$$

In a similar way, one obtains ($k = 0, 1, \ldots, n$):

$$\gamma^*_{n-k} = \exp\{A_{n-k}\} \cdot \Theta^*_{n-k}(E_{n-k}, F_{n-k}) \tag{68}$$

where

$$\Theta^*_{n-k}(E_{n-k}, F_{n-k}) = \min_{u_{n-k} \in \Omega(u_{n-k})} \Theta_{n-k}(E_{n-k}, F_{n-k}, u_{n-k}) \tag{69}$$

And so it is seen that it is only necessary to memorize the $\Theta_n^* - k$ functions of two variables, which can be accomplished without any considerable difficulties. In minimization, it is sufficient to verify the extreme values of $u_{n-k} = +1$.

Conclusion

The dual control theory may be extended in various directions. Thus, for example, its extension to purely discrete systems merits attention, in which each of the magnitudes can assume only one of the permissible levels.

The development of this theory makes it possible to clarify the principles involved in the optimal teaching of discrete automatic machines.

The theory described above pertains to the 'Beiesov' type, inasmuch as the assumption is made in it, that the *a priori* probability characteristics are known. However, the formulation of the dual control theory is likewise expedient for those cases where these characteristics are unknown. Such a formulation may be carried out either on the basis of the minimax principle or by the application of the idea of inductive probability.

At the present time, the most important problem for the immediate future is the development of approximate solution methods for dual control theory problems, the formulation of sub-optimal strategies, the determination of the numerical value of risk in practically optimal systems and its comparison with the value of risk in existing systems. Such a comparison will make it possible to clarify the extent of the gain that may be anticipated where we have a maximum degree of perfection in existing systems.

References

[1] BELLMAN, R. *Dynamic Programming*. 1960. Inoizdat

[2] PONTRIAGIN, L. S., BOLTIANSKII, V. G., GAMKRELIDZE, R. V. and MISHCHENKO, E. F. *Mathematical Theory of Optimal Processes*. 1961. Fizmatgiz

[3] CHANG, S. S. L. *Optimum Synthesis of Control Systems*. 1961

[4] FELDBAUM, A. A. *Computers in Automatic Systems*. 1959. Fizmatgiz

[5] PUGACHEV, V. S. *Theory of Random Functions and its Application to Automatic Control Problems*. 1960. Fizmatgiz

[6] LENING, D. KH. and BETTIN, R. G. *Random Processes in Automatic Control Problems*. 1958. Inoizdat

References 7–12 to be supplied

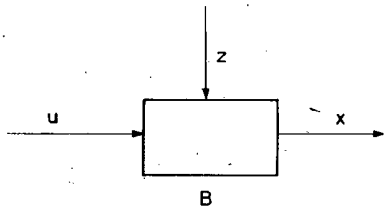


Figure 1

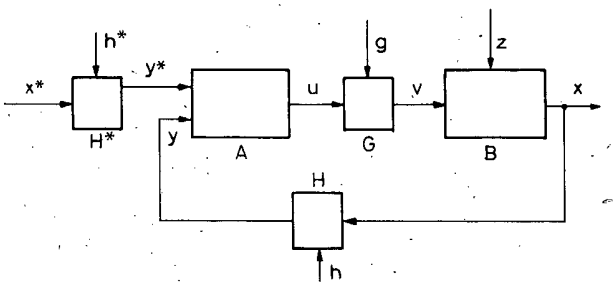


Figure 2

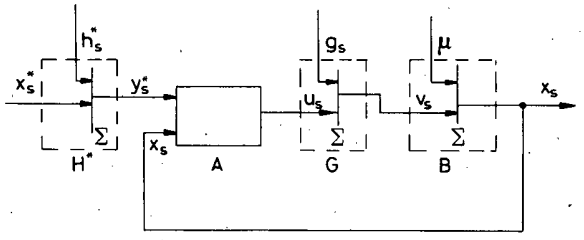


Figure 3

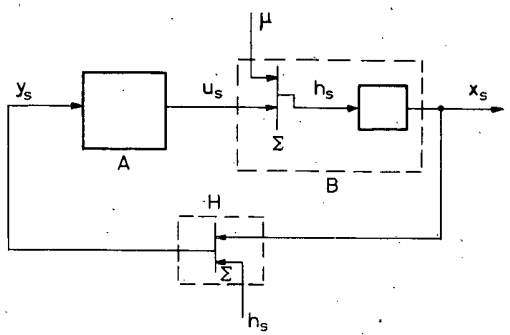


Figure 4

Fundamentals of the Theory of Non-linear Pulse Control Systems

Ya. Z. TSYPKIN

Introduction

The theory of linear pulse control systems has attained a high level of development and the main problems in the analysis and synthesis of such systems can be solved. However, with regard to non-linear pulse control systems, the theory is still in its initial stage. Up to the present time non-linear theory has been confined mostly to the investigation of periodic conditions. Yet periodic conditions are not operational conditions and the important problem still remains to ensure the stability of non-linear pulse control systems and to assess the 'quality' of stable processes. Attempts to employ the methods of investigating periodic conditions for estimating stability when the required periodic conditions are no longer present are often unjustified since the absence of a particular type of periodic condition is no guarantee that other forms of periodic or almost-periodic conditions are not present.

For solution of the stability problem it was quite natural to try to employ the ideas of Liapunov's second method which is widely used in the theory of continuous systems, in extending them to difference equations[1–5].

However, such an approach involves difficulties associated with the need to transform the equations of non-linear pulse systeme into their normal form, the arbitrariness of the selection of Liapunov functions and the impossibility of establishing any general properties of non-linear pulse control systems.

The approach to the problem in this paper is based on an idea which Popov[6, 7] used in the investigation of non-linear continuous control systems. The distinctive feature of this approach is that it is closely associated with such physical concepts as the frequency and, transient responses, and it provides the widest sufficient conditions of stability which can be obtained by all the Liapunov functions of the quadratic type. This approach greatly simplifies an assessment of the quality of processes in non-linear pulses control systems. It is possible to establish when the absence of periodic solutions guarantees stability and, finally, use may be made of methods similar to those employed in the investigation of linear pulse systems.

Statement of the Problem

A block diagram of a non-linear pulse control system is shown in *Figure 1*. It consists of a non-linear element in series with a linear pulse part LP which is an open linear pulse loop. The linear pulse part incorporates a pulse element for amplitide modulation of arbitrarily shaped pulses and a continuous part.

Letus suppose that the characteristic $\Phi(x)$ of the non-linear element satisfies the following conditions (*Figure 2*):

$$\text{(a)} \quad \Phi(0)=0$$

$$\text{(b)} \quad 0<\frac{\Phi(x)}{x}<k_0 \qquad (1)$$

$$\text{(c)} \quad \lim_{x\to\pm\infty} \Phi(x)\pm 0$$

which correspond to the fact that this characteristic belongs to the interval $(0, k_0)$.

The main problem is to determine the stability of the systems which are to be considered for any initial deviations and to determine the quality of behaviour in stable systems. Stability of this kind which is independent of the particular shape of the characteristic of the non-linear element and which satisfies the general conditions (1), is called generally absolute stability[8].

Equations of Non-linear Pulse control Systems

Let one suppose that the continuous part of the linear pulse part *LP* receives perturbations in the form of initial conditions with $n = 0$. One puts $f[n]$ for the response of this continuous part to partial conditions and applies it to the input of the non-linear pulse system (*Figure 1*). If the continuous part, and therefore the linear pulse part, is stable, then

$$\lim_{n\to\infty} f[n]=0 \qquad (2)$$

The equation of the pulse control system with respct to the error $x[n]$ can take either of two forms.

(*i*) With respect to original lattice functions:

$$x[n]=f[n]-\sum_{m=0}^{n} w[n-m]\Phi(x[m]) \qquad (3)$$

(*ii*) With respect to their transforms:

$$X^*(q)=F^*(q)-W^*(q)\,D\{\Phi(x[n])\} \qquad (4)$$

Here[9]

$$Z^*(q)=D\{z[n]\}=\sum_{n=0}^{\infty} e^{-qn} z[n] \qquad (5)$$

is the discrete Laplace transform (D transformation): $q=\sigma+j\bar{\omega}$ is a parameter of transformation; $\bar{\omega}=\omega T$ is the relative frequency: T is the repetition interval[9];

$$W^*(q)=D\{w[n]\} \qquad (6)$$

is the transfer function of the linear pulse part;

$$w[n]=w(\bar{t}),\quad \bar{t}=n \qquad (7)$$

is the impulse characteristic of the linear pulse part; $x[n]$, $f[n]$ are the lattice functions, which correspond to the error and the

reduced input: $X^*(q)$, $F^*(q)$ are their transforms and, finally, $\Phi(x)$ is the characteristic of the non-linear element.

For a stable linear pulse part

$$\lim_{n\to\infty} w[n]=0 \tag{8}$$

This implies that the corresponding transfer function $W^*(q)$ has no poles in the right-hand half-band $\mathrm{Re}\,q\geq 0$, $-\pi<\mathrm{Im}\,q\leq\pi$.

The Sufficient Condition of Absolute Stability

A pulse control system is absolutely stable relative to any perturbation $f[n]$ which satisfies the condition (2) if

$$\lim_{n\to\infty} x[n]=0 \tag{9}$$

In order to establish the fact of absolute stability, one estimates the solutions $x[n]$ of the equation with respect to the original functions.

By analogy with the ideas of Popov[6, 7], the auxiliary functions are now introduced

$$\varphi_N[n]=\begin{cases}\Phi(x[n]) & 0\leq n\leq N\\ 0 & n<0,\ n>N\end{cases} \tag{10}$$

and

$$\psi_N[n]=x_N[n]-\frac{1}{K}\varphi_N[n] \tag{11}$$

where

$$x_N[n]=f[n]-\sum_{m=0}^{n} w[n-m]\,\varphi_N[m] \tag{12}$$

It is obvious that for $0\leq n\leq N$

$$x_N[n]\equiv x[n]$$

where $x[n]$ is the solution of eqn (3).

Now the following expression is formed

$$\rho_N=\sum_{n=0}^{\infty}\varphi_N[n]\,\psi_N[n] \tag{13}$$

which, having regard to (10) and (11), is equal to

$$\rho_N=\sum_{n=0}^{\infty}\left(\Phi(x[n])\,x[n]-\frac{1}{K}\Phi^2(x[n])\right) \tag{14}$$

According to the Liapunov-Parseval equality[9] eqn (13) can also be represented as

$$\rho_N=\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_N^*(-j\bar{\omega})\,\psi_N^*(j\bar{\omega})\,d\bar{\omega} \tag{15}$$

where

$$\Phi_N^*(j\bar{\omega})=D\{\varphi_N[n]\}_{q=j\bar{\omega}} \tag{16}$$

and by virtue of (11) and (12)

$$\psi_N(j\bar{\omega})=D\{\psi_N[n]\}_{q=j\bar{\omega}}=F^*_{(j\bar{\omega})}-\left(W^*(j\bar{\omega})+\frac{1}{K}\right)\Phi_N^*(j\bar{\omega}) \tag{17}$$

These spectral functions exist if conditions (10) and (8) are fulfilled.

Substituting (16) and (17) into (15) and after simple transformations one gets

$$\rho_N=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sqrt{\mathrm{Re}\,\Pi^*(j\bar{\omega})}\,\Phi_N^*(j\bar{\omega})-\frac{F^*(j\bar{\omega})}{2\,\mathrm{Re}\,\Pi^*(j\bar{\omega})}\right|^2 d\bar{\omega}+\frac{1}{8\pi}\int_{-\pi}^{\pi}\frac{F^*(j\bar{\omega})}{\mathrm{Re}\,\Pi^*(j\omega)}d\bar{\omega} \tag{18}$$

where

$$\mathrm{Re}\,\Pi^*(j\bar{\omega})=\mathrm{Re}\,W^*(j\bar{\omega})+\frac{1}{K}>0 \tag{19}$$

The function

$$\Pi^*(j\bar{\omega})=W^*(j\bar{\omega})+\frac{1}{K}$$

which plays the main role, is called the analogue of the Popov function.

Since the first integral in (18) is negative, by discarding it, one obtains the inequality

$$\rho_N\leq\frac{1}{8\pi}\int_{-\pi}^{\pi}\frac{|F^*(j\bar{\omega})|^2}{\mathrm{Re}\,\Pi^*(j\bar{\omega})}d\bar{\omega}=C \tag{20}$$

By virtue of (19) the quantity c is positive: it is independent of N.

Substituting into the left-hand side of (20) the value of ϱ_N from (14) one obtains

$$\sum_{n=0}^{N}\Phi(x[n])\,x[n]\left(1-\frac{\Phi(x[n])}{Kx[n]}\right)\leq C \tag{21}$$

According to the condition (1a), the sum on the left-hand side of (21) is positive, moreover it is limited. The series which is formed from this sum as $N\to\infty$, therefore converges. Using the known theorem of the convergence of series with positive terms, one concludes that

$$\lim_{n\to\infty}\Phi\left(x[n]\,x[n]\left(1-\frac{\Phi(x[n])}{Kx[n]}\right)\right)=0$$

Hence, by virtue of the conditions (1), it follows that

$$\lim_{n\to\infty} x[n]=0 \tag{22}$$

Thus a pulse control system which has a stable pulse linear part and a non-linear characteristic $\Phi(x)$ and which satisfies the conditions (1), will be absolutely stable if the real part of the analogue of Popov's function is positive, i.e. if

$$\mathrm{Re}\,\Pi^*(j\bar{\omega})=\mathrm{Re}\,W^*(j\bar{\omega})+\frac{1}{K}>0 \tag{23}$$

The condition of stability (23) determines the magnitude of the interval $(0, k)$ which includes the non-linear characteristic $\Phi(x)$ for which the pulse system is absolutely stable. This condition is sufficient.

Frequency Criteria of Absolute Stability

To formulate the criteria of stability of a pulse control system one introduces the concept of a static gain of the non-linear element

$$S(x)=\frac{\Phi(x)}{x} \tag{24}$$

which is the slope of a straight line passing through the point of the non-linear characteristic for a specified value of x. The maximum S_{max} and the minimum S_{min} static gains are determined by the rays of a sector which is tangential to the characteristic (*Figure 3*). A non-linear pulse control system in which the non-linear element is replaced by a linear element with some fixed gain, k, is said to be a linearized pulse control system. For a linearized pulse system to be stable, by analogy with the Nyquist criterion[7], it is necessary and sufficient that the frequency characteristic of the linear pulse part LP should not embrace the ponts $-1/k, j0$. It will be said that a linearized system is obvtously stable if the frequency characteristic of the linear pulse part does not intersect the straight line $-1/k$. Then, according to the condition of stability (23), the frequency criterion of absolute stability of a non-linear pulse control system can be formulated in the following way. A non-linear pulse control system with its characteristic belonging to the interval $(0, k)$, will be absolutely stable if the linearized pulse system corresponding to it is obvtously stable or if the frequency characteristic $W^*(j\bar{\omega})$ of the linear pulse part does not intersect the straight line $-1/k$ (*Figure 4*).

The greatest value $k = k^0$ which determines the span of the interval (sector) in which the non-linear characteristic is located, is determined by drawing the vertical tangent to $W^*(j\bar{\omega})$. The difference $k - S_{max}$ characterizes the margin of stability.

The stability criterion of a pulse control system can also be formulated with referenec to the frequency characteristic $K^*(j\bar{\omega})$ of a closed linearized pulse control system. Selecting $k = k_0/2$; then

$$K^*(j\bar{\omega}) = \frac{\frac{k_0}{2} W^*(j\omega)}{1 + \frac{k_0}{2} W^*_{\perp}(j\bar{\omega})} \tag{25}$$

According to the usual constructions of the frequency characteristic of a closed loop from th e frequency characteristic of an open loop[9], for a obvtously stable linearized pulse control system if $k = k_2/2$, one has

$$|K^*(j\bar{\omega})| \leq 1 \tag{26}$$

Thus a non-linear pulse control system with its characteristic belonging to the interval $(0, k_0)$ will be absolutely stable if the frequency characteristic of the closed linearized pulse control system $K^*(\bar{\omega})$ with gain $k_0/2$ does not exceed unity in absolute value.

One Notes that the frequency criteria are also applicable in those cases when the continuous part contains delay elements or elements with distributed constants.

The frequency criteria of absolute stability can also be expressed in analytic form. The first criterion is closely related to the problem of Karatsodor, whilst the second criterion is closely associated with Shur's problem in the theory of analytic functions[10].

The analytic form of the criteria is considered in a special paper. One will not consider it here as, more over, the use of frequency criteria is the simplest way of elucidating various general properties of non-linear pulse control systems.

Generalization of the Stability Criteria

Non-linear pulse control system which contain a stable linear pulse part have been considered above. Now suppose that the linear pulse part is neutral or unstable. This implies that its transfer function $W^*(q)$ has poles on the imaginary axis, and in particular, at the origin or in the right-hand half-band $\mathrm{Re}\, q \geq 0, -\pi < \mathrm{Im}\, q \leq \pi$[9]. Since the determined sufficient conditions must hold for any non-linear characteristic which belongs to the interval $(0, k_0)$, they must also hold for a linear characteristic which belongs to this interval. But for sufficiently small gains z of this linear characteristic, a closed pulse control system will behave like an open pulse control system corresponding to the linear pulse part, i.e. it will be neutral or unstable. Therefore, for instances of a neutral or unstable linear pulse part it is necessary to impose additional limitations on the minimum static gain S_{min}. Let us elueidate these limitations. Given a proportional feedback with the coefficient z across the linear pulse part (*Figure 6*), one supposes that the structure of the linear pulse part is such that for a finite $z < S_{min}$ the closed linear pulse part is stable. The frequency criteria of stability are then applicable to this non-linear pulse control system, but the role of the frequency characteristic of the linear pulse part $W^*(j\bar{\omega})$ will now be played by the frequency characteristic of the closed pulse control system, which is a new linear pulse part equal to

$$W_e^*(j\bar{\omega}) = \frac{W^*(j\bar{\omega})}{1 + zW^*(j\bar{\omega})} \tag{27}$$

But the blockdiagram of a non-linear pulse control system [*Figure 6(a)*] can easily be converted to the form of *Figure 6(b)* where $f[n]$ is now the response of the closed pulse control system, and the non-linear characteristic is equal to

$$\Phi(x) + zx \tag{28}$$

However, since this characteristic must satisfy the conditions (1),

$$z < \frac{\Phi(x)}{x} < k \tag{29}$$

i. e. $$S_{min} > z$$

Thus the formulation of the frequency criterion remains unchanged. Only the characteristic of the non-linear element must now belong to the sector (z, k), and the frequency characteristic of the linear pulse part $W_e^*(j\bar{\omega})$ is determined by the expression (27).

One Notes that if the linear pulse part is neutral and its transfer function $W^*(q)$ has only one zero pole, whilst the rest of the poles have negative real parts, then z in eqn (27) can be arbitrarily small and for this case one has

$$W_e^*(j\bar{\omega}) \approx W^*(j\bar{\omega}) \tag{30}$$

i.e. in this case there is no need to construct $W_e^*(j\bar{\omega})$ from $W^*(j\bar{\omega})$ on the basis of the relation (27).

If the non-linear characteristic $\Phi(x)$ at $x \geq x^0$ goes outside the limits of the sector (z, k), which is usually the case for non-linear characteristics of the saturation type, the frequency criterion of stability guarantees stability with deviations of the error not exceeding x^0.

The frequency criteria of stability also hold for those cases when the non-linear characteristic (or gain of the linear pulse part) is a function of time n, if $\Phi(x, n)$ for any $n \geq n_0$ satisfies the conditions (1), i.e. if it belongs to the sector $(0, k_0)$ or in the case of a neutral or unstable linear pulse part belongs to the interval (z, k_0).

The Necessary and Sufficient Conditions of Absolute Stability for Some Non-linear Control Systems

Frequency criteria of absolute stability determine the sufficient conditions of absolute stability. It is obvious that in those cases when these sufficient conditions of absolute stability coincide with the necessary and sufficient condition of stability of linearized pulse control systems, they also become necessary conditions of absolute stability. Let us define the class of non-linear pulse control systems for which the conditions of absolute stability are necessary and sufficient. This problem was first posed by Aizerman[11], for continuous control systems, and slightly later by Letov[8]. The solution of this problem is of importance since it permits reduction of the investigation of the absolute stability of non-linear pulse control systems to the well-known investigation of the stability of linear pulse control systems.

It follows directly from the formulation of the frequency criterion that this class of non-linear pulse control systems includes those for which the obvtous stability of linearized pulse control systems coincides with their stability. The frequency characteristics of these latter pulse control systems $W^*(j\bar{\omega})$ [or $W_e^*(j\bar{\omega})$] must have the form shown in *Figure 7(a)* and *(b)*. The frequency criterion of absolute stability determines the necessary and sufficient conditions for all non-linear pulse control systems of the first order (with amplitude- or pulse width- or time-modulation), and also for non-linear pulse control systems of any order whose frequency characteristic $W^*(j\bar{\omega})$ has the largest real part in absolute value at the boundary frequency. It is worthwhile pointing out that for this class of system the absence of periodic conditions according to the improved method of harmonic balance[12], testifies to their stability. For digital automatic systems, as shown elsewhere[13], the determination of periodic conditions with a relative frequency $\bar{\omega} = \pi$ entails drawing a straight line with a slope $-1/W^*(j\bar{\omega})$ in the plane of the non-linear characteristic (*Figure 7*)*. If the maximum real part $W^*(j\bar{\omega})$ in absolute magnitude is attained for $\bar{\omega} = \pi$ (which always occurs for firstorder pulse control systems), the condition requiring the absence of aperiodic conditions with a relative frequency $\bar{\omega} = \pi$ coincides with the condition of absolute stability.

Estimation of the Degree of Stability

For the simplest estimate of the quality of the behaviour of a non-linear pulse control system, onewill use the concept of degrees of stability which characterizes the process damping speed.

For this purpose, instead of the auxiliary functions (10) and (11), the following functions are introduced.

$$\varphi_N[n]=\begin{cases}\Phi(x[n])\,e^{\delta n} & 0\le n\le N\\ 0 & n<0,\ n>N\end{cases} \tag{31}$$

and

$$\psi_N[n]=x_N[n]\,e^{\delta n}-\frac{1}{k}\varphi_N[n]\,e^{\delta n} \tag{32}$$

where $\delta > 0$ is some constant quantity.

Multiplying both sides of (12) by $e^{\delta n}$, there is obtained

$$x_N[n]\,e^{\delta n}=f[n]\,e^{\delta n}-\sum_{m=0}^{n} w[n-m]\,e^{\delta(n-m)}\psi_N[m]\,e^{\delta m} \tag{33}$$

* The author points out that in a previous paper[12] he has given an erroneous slope.

Remarking that according to the shift theorem[9]

$$D\{z[n]\,e^{\delta n}\}_{q=j\bar{\omega}}=Z^*(-\delta+j\bar{\omega}) \tag{34}$$

and following the same discussion as in the establishment of the condition of absolute stability, the conclusion is reached that

$$\lim\ x[n]\,e^{\delta n}=0 \tag{35}$$

if the real part of the analogue of the shifted function of Popov is positive, i.e. if

$$\operatorname{Re}\Pi^*(-\delta+j\bar{\omega})=\operatorname{Re}W^*(-\delta+j\bar{\omega})+\frac{1}{k}>0 \tag{36}$$

As will be seen from eqn (35), the rate of damping is determined here by the quantity δ. The determination of the conditions for which non-linear pulse control systems have a specified degree of stability, δ_0, thus entails the use of the frequency criterion of stability and its application to the shifted frequency characteristic

$$W^*(-\delta_0+j\bar{\omega}) \tag{37}$$

or

$$W_e^*(-\delta+j\bar{\omega})=\frac{W^*(-\delta_0+j\bar{\omega})}{1+zW^*(-\delta_0+j\bar{\omega})} \tag{38}$$

for a fixed value δ_0 (*Figure 9*).

Since the poles of the transfer function $W^*(-\delta+q)$ depend on δ and with increase of δ are shifted in the direction of the right-hand half-band, the greated value of $\delta = \delta_{\max}$ is attained for a value

$$z_0 \le S_{\min} \tag{39}$$

which still ensures stability of a closed linear pulse part. Thus the increase of δ is possible until the poles $W^*(-\delta+q)$ are located at the origin or on the imaginary axis. With an increase of δ the quantity k^0 usually decreases, whilst z increases. Therefore, the less the difference $S_{\max} - S_{\min}$, the more attainable is a large degree of stability. This estimate is also applicable to non-linear pulse control systems in which he characteristic of the non-linear element depends also on time.

The Overall Quadratic Estimate

Another important estimate of the quality of behaviour is the overall quadratic estimate of the output of a non-linear element

$$I_2=\sum_{n=0}^{\infty}\Phi^2(x[n]) \tag{40}$$

To determine the upper boundary of this estimate, one will avail one-self of Popov's ideas[14]. Consider the inequality (21) for $N = \infty$, representing it in the form

$$\rho_\infty=\sum_{n=0}^{\infty}\Phi^2(x[n])\left(\frac{x[n]}{\Phi(x[n])}-\frac{1}{K}\right)\le C \tag{41}$$

Since

$$\frac{x}{\Phi(x)}>\frac{1}{S_{\max}}$$

the inequality (41) can be strengthened and

$$\left(\frac{1}{S_{\max}}-\frac{1}{K}\right)\sum_{n=0}^{\infty}\Phi^2(x[n])\le C \tag{42}$$

Taking into account the notation of (40) and (20), from (42) one gets

$$I_2 \leq \frac{k S_{max}}{k - S_{max}} \frac{1}{8\pi} \int_{-\pi}^{\pi} \frac{|F^*(j\bar{\omega})|^2}{\mathrm{Re}\,\Pi^*(j\bar{\omega})} d\bar{\omega} \quad (43)$$

where

$$\mathrm{Re}\,\Pi^*(j\bar{\omega}) = \mathrm{Re}\,W^*(j\bar{\omega}) + \frac{1}{k} \geq \frac{1}{k} - \frac{1}{k_0} > 0 \quad (44)$$

Replacing $\mathrm{Re}\,\Pi^*(j\bar{\omega})$ in eqn (43) by its maximum value, one finally gets

$$I_2 \leq \frac{k_0 k^2 S_{max}}{(k - S_{max})(k_0 - k)} \frac{1}{8\pi} \int_{-\pi}^{\pi} |F^*(j\bar{\omega})|^2 d\bar{\omega} \quad (45)$$

The right-hand side of inequality (44) contains an undetermined parameter k; here (*Figure 3*)

$$S_{max} < k < k_0 \quad (46)$$

This is so selected that the coefficient is minimum for the integral (45). It can be shown without difficulty that in this case

$$k = \frac{k_0 - S_{max}}{2 k_0 S_{max}} \quad (47)$$

and therefore finally get

$$I_2 \leq \frac{k_0^2 S_{max}^2}{(k_0 - S_{max})} \frac{1}{2\pi} \int_{-\pi}^{\pi} |F^*(j\bar{\omega})|^2 d\bar{\omega} \quad (48)$$

But according to the Liapunov-Parseval[9] equality,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |F^*(j\bar{\omega})|^2 d\bar{\omega} = \sum_{n=0}^{\infty} f^2[n] \quad (49)$$

Therefore eqn (48) can also be represented as

$$I_2 \leq \frac{k_0^2 S_{max}^2}{(k_0 - S_{max})} \sum_{n-0}^{\infty} f^2[n] \quad (50)$$

It follows from (50) that the upper boundary of the overall quadratic estimate is determined by the sum of the squares of the discrete responses of the linear impulse part to the applied inputs. If the linear impulse part receives an input $f_1[n]$ which decreases with time, then

$$f[n] = \sum_{m=0}^{n} w[n-m] f_1[m]$$

and this implies that

$$F^*(j\bar{\omega}) = D\{f[n]\}_{q=j\bar{\omega}} = \vartheta\{W^*(q) F_1^*(q)\}_{q=j\bar{\omega}}$$

The computation of the right-hand sides of (48) or (50) is carried out analytically or graphically by known rules[9]. The upper boundary of the overall quadratic error is less, other things being equal, the greater the margin of stability $k_0 - S_{max}$. This estimate is also applicable when the characteristic of a non-linear element depends also on time.

Conclusion

This approach to the problem makes it comparatively simple by the concepts of the linear theory of pulse systems to determine the region of absolute stability of non-linear pulse control systems and to estimate indices of the quality of processes (the degree of stability and the overall quadratic estimate). The fact that the stability and estimates of indices of process quality are independent of the actual shape of the characteristic of the non-linear element, provided only that this characteristic belongs to the specified sector, makes it possible to ensure values of estimates of the indices of quality for variation of the characteristic of the non-linear element or of the parameters of the linear pulse part which also lead to a change in the boundaries of the sector (z, k). In some cases it is therefore no longer necessary to use special additional self-adjusting circuits which complicate non-linear pulse control systems.

In this connexion it is extremely important to determine the structure of non-linear pulse control systems, the sensitivity[15] of which is low in relation to variations of the non-linear characteristic and to the parameters of the linear part. For this purpose use may be made of the results of investigations into the sensitivity of linear pulse control systems.

Generalization of the method of investigating non-linear pulse control systems to pulse control systems which contain a linear pulse part with time-variable parameters, and several non-linear elements, widens the range of problems which can be solved and, in particular, makes it possible to investigate non-linear pulse control systems in wich pulse-width, pulse-phase and pulse-frequency modulation is provided.

References

1 BROMBERG, P. V. *Stability and self-oscillation of sampled-data control systems.* 1953. Moscow

2 HAHN, W. Über die Anwendung der Methode von Liapunov auf Differenzengleichungen. *Mat. Annalen* Bd. 136 (1958) 430–441

3 HAHN, W. *Theorie und Anwendung der direkten Methode von Liapunov.* 1959. Springer Verlag

4 KALMAN, R. E. and BERTRAM, J. E. Control system analysis and design via the second method of Liapunov: II. Discrete Time Systems. *Trans. ASME, J. of Basic Engineering*, S.D., 82 No. 3 (1960) 371–400

5 BERTRAM, J. E. *The direct method of Liapunov in the analysis and design of discrete time control systems.* Work session in Liapunov's second method. Ed. L. K. Kazda, University of Michigan (1960) 79–104

6 POPOV, V. M. Criterii de stabilitate pontru sistemele nolineare de reglage avtomata po utilizarea transformatici Laplace. *Studii si Cecretari de Energetica.* An. 9, 1 (1959) 119–135

7 POPOV, V. M. Concerning the absolute stability of non-linear control systems (Ob absolyutnoi ustoichivostin elineinykh sistem avtomaticheskogo regulirovaniya). *Avtomat. i telemekh* 22, No. 8 (1961) 961–979

8 LETOV, A. M. *The stability of non-linear control systems.* 1955. Moscow; Gostekhizdat

9 TSYPKIN, Ya. Z. *Theory of pulse systems.* 1958. Moscow; Fizmatgiz

10 AKHIEZER, N. I. *The classical problem of moments.* 1961. Moscow; Fizmatgiz

11 AIZERMAN, M. A. Concerning a problem which touches on the stability 'in the large' of dynamic systems. *Uspekhi mat. nauk* 4, 4 (32) (1949) 186–188

12 TSYPKIN, Ya. Z. Periodic modes in non-linear pulse-type control systems. *Trud. Tashkent Polytechnical Institute* (New series) *Energetika*, 20 (1961) 184–195

13 TSYPKIN, Ya. Z. Elements of the theory of digital control systems. *Automatic and Remote Control.* 1960. London; Butterworths. *Akad. Nauk SSSR*, 1961. 2 (1961) 63–79

14 POPOV, V. M. A criterion of the quality of non-linear control systems. *Automatic and Remote Control.* 1960. Butterworths; *Akad. Nauk SSSR* 1 (1961) 404–441

15 HOROWITZ, I. M. The sensitivity problem in pulse feedback systems. *IRE Trans. Automatic Control* AC-6, No. 3 (1961) 251–260

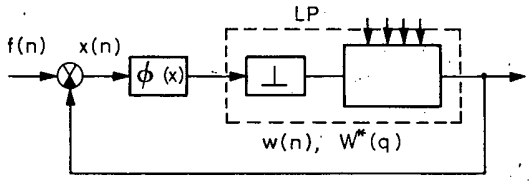


Figure 1

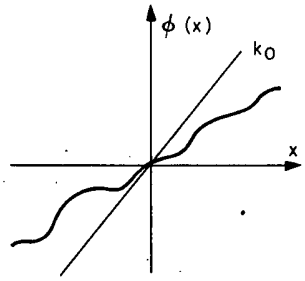


Figure 2

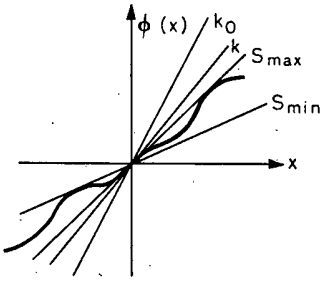


Figure 3

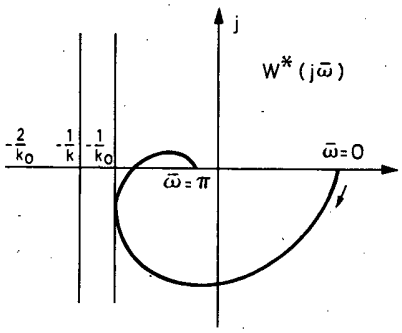


Figure 4

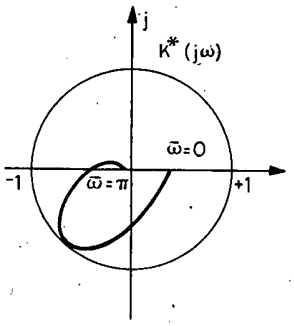


Figure 5

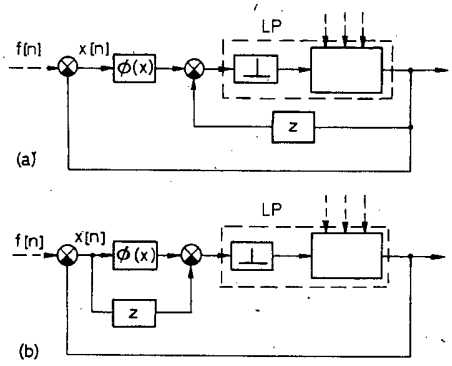


Figure 6

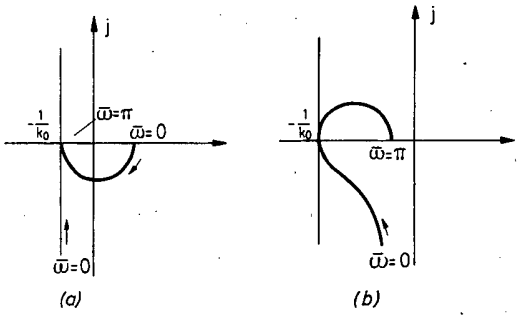


Figure 7

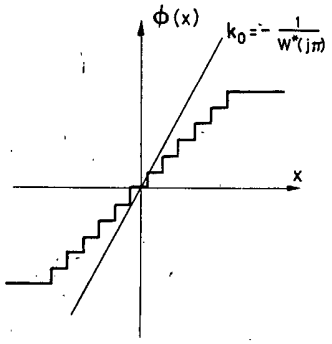


Figure 8

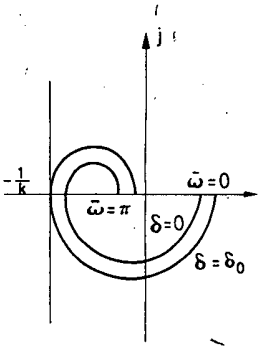


Figure 9

Synthesis of Optimum Sampled-data Systems

L. N. VOLGIN

Introduction

The transition to the scientific design of compound automatic complexes and the resulting increase in the calculation difficulties demand the finding of new ways for the formalization of solutions and simplification of calculation. The creation of the methods of linear, non-linear and dynamic programming should be regarded as considerable achievements in this field. The method of polynomial equations used below, the efficiency of which was demonstrated on a number of problems of automatic control, may be added to this number of methods. The polynomial equations consist of a variety of diophant equations, the specific methods of solution of which are easily programmed for digital computers. The development of the operator method of analysis for the linear pulse systems, of the discrete Laplace transformation, or z-transformation (see Tsypkin[1], Gurevich[2], Zadeh and Ragazzini[3], and others), and the emergence of a large number of different methods for the synthesis of the optimum linear pulse systems (the works of Tsypkin[4–6], Bergen and Ragazzini[7], Chang[8–10], Jury[11], Bertram[12], Potapov[13], Krasovskii[14], Perov[15], and others) were the reason for the creation of the method of polynomial equations.

At present the theory of optimum pulse systems for the control of linear plants lies at the foundation of design of self-optimizing systems, which contain digital computers. In these systems the automatic linearization of equations for the plant during the operation of the system is achieved on the basis of principles described in the works of Kalman[16], Bigelow, and Ruge[17], and others. Thus, the theory of optimum linear pulse systems develops into the theory of an extensive class of self-optimizing pulse systems of control, adaptable to the changing characteristics of the controlled plant and to the parameters of external signals.

The basic difficulties which arise in the design of the systems containing an optimizing model for the medium are associated with the violation of the conditions of 'approximation' of simulation, which require a continuous relationship between the quality of control and the change in the parameters of the plant being simulated. Some of the above-mentioned authors touch upon the questions of control for the plants with negative dynamic properties, during the compensation of which the violation of 'approximation' is possible. The criteria of approximation found under these conditions, which do not allow the contraction of the zero and poles of the transfer function of the plant for the individual structures of automatic systems, served as the starting point for the search of analytical conditions of approximation, suitable for any structures. The investigation of the conditions of approximation for automatic systems showed that they are closely connected with the conditions of stability, and that the distinctive feature of these conditions is based on the distinctive ideas about the 'coordinate' and 'parameter'. The conditions found below, which combine the conditions of stability and approximation, are called *the efficiency conditions*, since the term 'efficiency' literally reflects the essence of the considered phenomenon. From the analytical conditions of efficiency for different structures of automatic systems emerge different criteria for efficiency. On the basis of these criteria it is possible to conclude that the criteria of stability adopted at present are inadequate for the synthesis of efficient systems. The attempts to solve the problems of synthesis for automatic systems often encountered in literature, inaccurate on the whole or having a very limited field of application, are explained by this. The author has shown[18–20] that the polynomial equations represent a mathematical tool which is adequate for the problem of synthesis for efficient automatic pulse systems. A systematic treatment of the method of polynomial equations is contained in the author's monograph[20]. In the given paper a derivation of the analytical conditions of efficiency is given, and a brief survey made of the problems of automatic control, solvable by means of polynomial equations.

Denotations and Terms Used

1. Symbol z is used for transformations, where z is the delay operator for a single cycle.

2. The systems and signals are represented by the rational real functions of z of the form $F = A/B$, where A and B are polynomials of z.

3. The factorization of functions F with reference to contour $\Gamma\{|Z| = 1\}$ gives the real functions F^+ and F^-; $F = F^+F^-$, where the sign F^+ denotes the absence of zeros and poles of the function in the region $D^-\{|Z| \leq 1\}$, and the sign F^- denotes their absence in the region $D^+\{|Z| > 1\}$.

4. The separation of functions F with respect to contour Γ gives the real functions $F = F_+ + F_-$, where the sign F_+ denotes the absence of poles of the function in the region D^-, and the sign F_- denotes their absence in the region D^+.

5. The representaion (the transfer function) of the controlled plants will be made by $G = P/Q$, where P and Q are polynomials of z; the representation (the programme) of the pulse unit will be made by $W = C/D$, where C and D are polynomials of z; the representation of the pulse system as a whole will be made by H, and the representations for the input and output are equal to X and Y respectively.

Analytical Conditions for the Efficiency of Pulse Systems

By considering the mathematical model of an actual physical system, one is deliberately making a differentiation between the 'coordinates' of the system, the changes in which are reflected by the given model, and its 'parameters' which are determined

as fixed numbers which, in the given model, form the basis for calculations. However, the practice of construction of automatic systems shows that the uncontrollable discrepancies between the calculated and the actual parameters may be the cause of profound disparity in the calculated and the actual behaviour of the system. The failure to take this fact into account will sometimes lead to the construction of inefficient systems. The majority of automatic systems (the systems of stabilization and programme control, the computer and the reproduction systems, the systems for transmission and processing of data) require a continuous relationship for the behaviour and the small changes in external conditions, which are expressed in the change of input coordinates and parameters of the system. The conditions for which a continuous relationship between the coordinates of the system is observed are the conditions of stability. The conditions for which a continuous relationship between the behaviour of the system and the deviations of its parameters from the calculated values, which are assumed to be constant in a given model, is observed, are the conditions of approximation of simulation. The general condition of efficiency for an automatic system, constructed on the basis of a definite calculated model, which unites the conditions mentioned, may be formulated as follows. *With small variations in the input coordinates and parameters of the system the variations in the output coordinates should be small.*

Let us find the analytical conditions for the efficiency of an automatic pulse system, with a single input and a single output coordinate, described by the following difference equation:

$$\mathscr{F}(x_i, x_{i-1}, \ldots, x_{i-n}, y_i, y_{i-1}, \ldots, y_{i-m}) = 0 \qquad (1)$$

where $\mathscr{F}$ is the continuous function differentiable with respect to all arguments, i is the discrete time, and n and m are the corresponding number of stored values x at the input and y at the output. At the foundation of calculation of the system lies the linear model, obtainable by means of linearization of the equation of the system in the vicinity of the current 'operating point':

$$\sum_{k=0}^{n} \left(\frac{\partial \mathscr{F}}{\partial x_{i-k}}\right)_0 x_{i-k} + \sum_{k=0}^{m} \left(\frac{\partial \mathscr{F}}{\partial y_{i-k}}\right) y_{i-k} = 0 \qquad (2)$$

The numbers

$$a_k = \left(\frac{\partial \mathscr{F}}{\partial x_{i-k}}\right)_0; \quad b_k = -\left(\frac{\partial \mathscr{F}}{\partial y_{i-k}}\right)_0 \qquad (3)$$

which do not depend on index i over the interval of time under consideration, represent the equivalent parameters of the linear model.

Using z-transformation of number sequences[20], the equation for the linear model (2) may be written in the form:

$$Y = HX \qquad (4)$$

where H is the representation of the model, which is the rational function

$$H = \frac{A}{B}, A = \sum_{k=0}^{n} a_k z^k, B = \sum_{k=0}^{m} b_k z^k \qquad (5)$$

The representation of a real system, the parameters of which change in relation to time and coordinates, but sometimes also in an unexpected form, differs from the representation of its model by the variations $\delta H, \delta^2 H, \delta^3 H, \ldots$, which must satisfy the general condition for the efficiency of the system.

By varying the relation (4), the corresponding variations for the output of the system are obtained:

$$\begin{aligned} \delta Y &= H \cdot \delta X + \delta H \cdot X \\ \delta^2 Y &= H \cdot \delta^2 X + 2\delta H \cdot \delta X + \delta^2 H \cdot X \end{aligned} \qquad (6)$$

The conditions under which the variations in the output coordinate remain small have the form:

$$(\delta Y)_- = 0; \ (\delta^2 Y)_- = 0; \ (\delta^3 Y)_- = 0, \ldots \qquad (7)$$

By separating the right sides of expressions (6) the analytical conditions for the efficiency of the pulse system are obtained:

$$H_- = 0; \ (\delta H)_- = 0; \ (\delta^2 H) = 0, \ldots \qquad (8)$$

in which case the first of these conditions is the usual condition of stability, whereas the last are the conditions of 'approximation' of simulation. The necessity for taking into account the large variations is caused by the fact that as regards the parameters of the system its representation is a non-linear function. It is possible to construct an example where the violation of the efficiency is caused as much as is desired by a high variation[20]. However, in practice, mostly violations of the first two conditions of efficiency are encountered.

Criteria for the Efficiency of the Basic Structures of the Automatic Pulse Systems

The method of combining the controlled plants and the computing units is called the structural system of control. The simplest pulse systems of automatic control contain a single computing unit with representation (programme) W and a single controlled plant with representation G. To each structure of the system of control corresponds a definite function H, which depends rationally on W and G:

$$H = H(W, G) \qquad (9)$$

which is called the representation of the system. For each structure of control there is a definite class of permissible functions H, which may be realized in the system by the choice of different control programmes W, remaining at the same time within the limits of conditions of efficiency. The structures, which permit the realization of arbitrary functions H are called *the ideal structures*. The structures which do not have even a single permissible function are called *the inefficient structures*. From the point of view of the condition of stability only the stable functions of type H_+ are the permissible functions. However, if it is necessary to realize an unstable function, then the condition of stability may be discarded by limiting oneself to the fulfilment of the conditions of approximation.

By taking into account the variations in the representation of the controlled plant, simulated by function G, the conditions of efficiency (8) applied to system (9) may be written in the form:

$$H_- = 0; \left(\frac{\partial H}{\partial G} \cdot \delta G\right)_- = 0; \left(\frac{\partial^2 H}{\partial G^2} \cdot \delta^2 G\right)_- = 0; \ldots \qquad (10)$$

The functions $H, \partial H/\partial G, \partial^2 H/\partial G^2, \ldots$, derived by differentiation of (9), depend on W and G. In synthesis of systems for the automatic control of the programme of the computing unit, W is chosen in relation to the representation of plant G:

$$W = W(G) \tag{11}$$

The verification of the synthesized systems for efficiency is made by the substitution of this relationship in the expression (10) after carrying out the operations of differentiation in them.

In a general case the pulse systems of automatic control contain several controlled plants and computing units, which are connected up into a single structure. These systems may have several inputt and outputs. The verification of the conditions for efficiency should be carried out in this case by the variation of all the output coordinates for the variation in the representations of all the controlled plant.

The compensation for the negative dynamic properties of the controlled plant, by means of the computing unit having the same negative dynamic properties, is the cause of violation of the conditions of efficiency of pulse systems of automatic control. Namely, such a compensation takes place, for example, during the trivial recalculation of the programme for the computing unit W for a simple closed system, the representation of which is:

$$H = \frac{WG}{1+WG} \tag{12}$$

by the formula:

$$W = \frac{1}{G} \cdot \frac{H}{1-H} \tag{13}$$

by proceeding from the initial function H, which is chosen without taking into account the conditions of efficiency.

This assumption will be proved. By carrying out the factorization of the representation of the plant it is obtained that:

$$G = G^+ G^- \tag{14}$$

Functions G^+ and G^-, equal to:

$$G^+ = P^+/Q^+; \quad G^- = P^-/Q^- \tag{15}$$

are the positive and the negative portions of the representation of the plant.

The *positive* plant, which has representation G^+, is characterized by the following dynamic properties: stability, instantaneousness of reaction, and smoothness of transition process. The *negative* plant, which has representation G^-, displays negative dynamic properties: instability, retardation of reaction, and sudden ejections in transition process.

By modifying formula (12) one obtains:

$$\delta H = \frac{W}{(1+WG)^2} \cdot \delta G; \quad \delta^2 H = -\frac{2W^2}{(1+WG)^3} \cdot \delta^2 G; \ldots \tag{16}$$

First of all, conditions will be found under which the closed system is ideal, i.e. capable of reproducing the arbitrary function H. The corresponding programme for the calculating unit is chosen in accordance with formula (13). By substituting this formula in (16) one obtains:

$$\delta H = H(1-H)\frac{\partial G}{G}; \delta^2 H = -2H^2(1-H)\frac{\partial^2 G}{G^2}; \ldots$$

or

$$\delta H = H(1-H)\left(\frac{\delta P}{P} - \frac{\delta Q}{Q}\right);$$

$$\delta^2 H = -2H^2(1-H)\left(\frac{Q\delta^2 p}{P^2} - \frac{2\delta P \cdot \delta Q}{P^2} + \frac{2\delta^2 Q}{PQ}\right); \ldots \tag{17}$$

The conditions of efficiency (8) require that $P^- = Q^- = 1$. Thus, the closed system is ideal only in that case when the plant is positive. In the case of the plant with negative dynamic properties the function H is not reliazable because of the violation of the conditions of approximation.

It will be shown that the closed automatic system is efficient for any controlled plant, under which conditions the class of permissible functions of this system is equal to

$$H = P^- \theta F_+ \tag{18}$$

where F_+ is the arbitrary stable rational function of the form:

$$F_+ = A/B^+ \tag{19}$$

and θ is the polynomial which satisfies the polynomial equation in respect of the unknown polynomials θ and Π:

$$AP^-\theta + Q^-\Pi = B^+ \tag{20}$$

The corresponding programme of control has the form:

$$W = \frac{AQ^+\theta}{P+\Pi} \tag{21}$$

It will be verified whether the conditions of efficiency are fulfilled. By substituting (21) in (16) and by taking into account (20) one obtains:

$$\delta H = \frac{A\theta\Pi}{(B^+)^2} \cdot \frac{Q\delta P - P\delta Q}{P^+Q^+};$$

$$\delta^2 H = \frac{2A^2\theta^2\Pi}{(B^+)^3} \cdot \frac{Q^2\delta^2 P - 2Q\delta P\delta Q + 2P\delta^2 Q}{(P^+)^2 Q^+}; \ldots$$

The conditions of efficiency are fulfilled for any values of G. In the case of a stable controlled plant the polynomial θ, as follows from the polynomial eqn (20), becomes arbitrary, and the class of permissible functions is extended to

$$H = P^- F_+ \tag{22}$$

Thus, one proves the criterion for the efficiency of a closed system, which requires in addition to the fulfilment of the usual criterion of stability, that the programme of the computing unit does not shorten polynomials P^- and Q^-*.

Using the analytical conditions of efficiency, it is possible to derive the criterion of efficiency for any structures of automatic systems. By means of these conditions it is easy to prove, for example, the following well-known propositions:

(1) The systems on the limit of stability are inefficient.

(2) The open systems of control are efficient only for the stable plants.

* Applicable to the stable plants the criterion of non-contraction P^- was, for the first time, introduced in the work of Bergen and Ragazzini[7].

(3) The ideal structures of control for the plants having negative dynamic properties do not exist.

(4) The parallel system of control is ideal for stable plants; the sequential (cascade) system of control is ideal only for positive plants.

In view of the non-existence of ideal structures of control for the arbitrary plants, the criterion of efficiency of the automatic system more rigid than the criterion of stability. Only for the positive plants are the general criteria of stability of the linear systems adequate.

In order not to violate the conditions of efficiency, the optimum function H of the system should be sought for in the class of permissible functions. The wider the class of permissible functions for a given structure, the higher the quality of the optimum system, remaining conditions being equal. Therefore, in the synthesis of a system of control for a given plant, a structure of control having as wide a class of permissible functions as possible, a structure close to an ideal one, should be chosen.

The Use of Polynomial Equations in the Synthesis of Optimum Pulse Systems

It has been established that the classes of permissible functions for the pulse systems are expressed in terms of polynomial equations. In the author's work[18–20], it was shown that the synthesis of optimum pulse systems of control for the linear plants based on a number of basic criteria may be made entirely by means of polynomial equations. The finding of the optimum programme of control is, as a rule, reduced to the solution of a system of polynomial equations. The computation methods for the solution of a system of polynomial equations applicable to the use of digital computers have also been developed and their advantage over the ordinary methods in the synthesis of controlled programmes for the plants of a high order with complex correlational relationships was proved. By means of the polynomial equations, a number of new problems of automatic control, in particular for the unstable controlled plants, was solved. The basic problems for the synthesis of pulse systems and their solutions, obtained by the method of polynomial equations, omitting the proofs because of the lack of space, are now enumerated.

The problem of synthesis of the pulse system with the minimum transient period for a given input action:

$$X=R/S \tag{23}$$

where R and S are the polynomials of z, is reduced to the solution of the following polynomial equation:

$$P^{-}\theta+SQ^{-}\Pi=R \tag{24}$$

in respect of unknown polynomials θ and Π. The corresponding controlling programme is equal to:

$$W=\frac{Q^{+}\theta}{P+S\Pi} \tag{25}$$

The representation of the transient process has the form:

$$E=Q^{-}\Pi \tag{26}$$

The minimum duration of the transient process, which ensures the fulfilment of the conditions of efficiency, from the number of cycles, is equal to the sum of powers of polynomials P^{-} and Q^{-}.

With the limitation for the module of the controlling action:

$$|u_i|\leq r \quad (i=0,1,2,\ldots) \tag{27}$$

the corresponding problem is reduced to the finding of a non-minimum solution of the polynomial equation, which is found by special computing methods. The modification of the polynomial equation (24) leads to the derivation of a system which has no pulses.

The problem of synthesis based on the criterion of the minimum of the total quadratic error:

$$\mathscr{J}=\sum_{i=0}^{\infty}e_i^2=\frac{1}{2\pi j}\oint_{\Gamma}E(z)E(z^{-1})\frac{dz}{z} \tag{28}$$

is reduced to the solution of the system consisting of two polynomial equations:

$$\left.\begin{aligned}P^{-}\theta+Q^{-}\Pi&=I^{+}\tilde{P}^{-}\tilde{Q}^{-}\\ P^{-}\theta+U^{+}\phi&=I^{+}\tilde{P}^{-}\tilde{Q}^{-}\end{aligned}\right\}^{*} \tag{29}$$

in respect of the unknown polynomials θ, Π and ϕ. The polynomials I and U are the numerator and denominator of function $X(z)\,X(z^{-1})$. The corresponding controlling programme is equal to

$$W=\frac{Q^{+}\theta}{P^{+}\Pi} \tag{30}$$

The calculation of the quadratic error may also be made by means of the polynomial equation[20].

The problems of synthesis of the optimum pulse systems of automatic control and of processing of data for the random input signals, by taking into account the universal nature and the prevalence of quadratic dispersion criteria, represent the most favourable field for the application of polynomial equations. The general problem of synthesis of a pulse system, optimum according to the criterion of dispersion of the error for finite time of transition into the unshifted state is reduced to the solution of a system consisting of three polynomial equations, one of which secures the efficiency of the synthesized system, the second, the finiteness of the settling time and the third, the minimization of dispersion of the error. The solution of this general problem determines the solutions of the numerous particular problems of extrapolation, filtration, differentiation and integration of random processes by means of pulse computing units. The optimization of the pulse systems, by arbitrary criteria of quality, is reduced to the combination of the method of polynomial equations and the general methods of mathematical programming. By means of the theory of polynomial equations it is possible to synthetize the most economic programmes for the processing of data by the method of least squares. The obtained results show that the polynomial equations represent a suitable mathematical tool for the programming of many procedures of computer mathematics and of mathematical statistics, which are widely used in the self-optimizing systems of automatic control.

* Polynomial with the reversed order for the sequence of coefficients is denoted by symbol $\tilde{A}$.

Conclusions

The conditions of efficiency, formulated in this paper, limit the possibility of change in the dynamic properties of controlled media by means of pulse computing units. Under these conditions the worst properties of the plant—instability, retardation, fluctuation—are shown to be the most difficult to overcome. The limits of the accuracy of control for the dynamic plants by means of the pulse computing units whilst being wider than for the units of the continuous type, are, however, not limitless. Physically, this means that the inertia of the plants cannot be completely overcome. The problem of the theory of automatic control lies in the further clarification of the limits of possible accuracy of control, and the realization of these possibilities through the design of the most perfect controlling machines. It is hoped that the future development of polynomial equations will prove to be one of the important aids in the solution of this problem.

References

[1] TSYPKIN, Ya. Z. *Theory of Pulse Systems.* (Monogr.) 1958. Moscow; Fizmatgiz

[2] JAMES, H., NICHOLS, N and PHILLIPS, R. *Theory of Cascade (sequential) Systems.* Eds, (Monogr.) IL, 1951, Chapter U

[3] ZADEH, L. A. and RAGAZZINI, J. R. The analysis of pulse systems. *Trans. Amer. Inst. elect. Engrs* 71 Pt II (1952)

[4] TSYPKIN, Ya. Z. Design of a system for automatic control under stationary incidental actions. *Automat. Telemech., Moscow* 4 (1953)

[5] TSYPKIN, Ya. Z. Some questions relating to the synthesis of automatic pulse systems. *Automatika* 1 (1958)

[6] TSYPKIN, Ya. Z. Optimum processes in automatic pulse systems. *Izv. AN SSSR, OTN, energet. avtomat.* 4 (1960)

[7] BERGEN, A. R. and RAGAZZINI, J. R. Sample-data processing techniques for feedback control systems. *Trans. Amer. Inst. elect. Engrs* 73 Pt II (1954)

[8] CHANG, S. S. L. Statistical design theory for strictly digital sampled-data systems. *Trans. Amer. Inst. elect. Engrs* 76 Pt I (1957)

[9] CHANG, S. S. L. Statistical design theory for digital-controlled continuous systems. *Trans. Amer. Inst. elect. Engrs* 77 Pt II (1958)

[10] CHANG, S. S. L. *Syntesis of Optimum Control Systems.* 1961. New York; [redacted]

[11] JURY, E. I. *Sampled-data Control Systems* 1958. New York; [redacted]

[12] BERTRAM, J. E. Factors in the design of digital controllers for sampled-data feedback systems. *Trans. Amer. Inst. elect. Engrs* 75 Pt II (1956)

[13] POTAPOV, M. D. The problem of terminal time of control and peculiarities of synthesis of some systems of automatic control. *Trudy VVIA im. N.E. Zhukovskogo,* 1959

[14] KRASOVSKII, A. A. Synthesis of the correcting pulse units of the cascade systems. *Automat. Telemech., Moscow* 6 (1959)

[15] PEROV, V. P. *Statistical Synthesis of Pulse systems.* (Monogr.) Sovetskoe radio, 1959

[16] KALMAN, R. E. Design of self-optimizing control systems. *Trans. Amer. Soc. mech. Engrs* 80, 2 (1958)

[17] BIGELOW, S. C. and RUGE, H. An adaptive system using periodic estimation of the pulse transfer function. *I.R.E. Conv. Rec.* IV (1961)

[18] VOLGIN, L. N. Method of synthesis of linear pulse systems for automatic control based on dynamic criteria. *Automat. Telemech., Moscow* 20 No. 10 (1959)

[19] VOLGIN, L. N. and SMOLYAR, L. I. The correction of cascade systems by means of certain calculating units. *Automat. Telemech., Moscow* 21 No. 8 (1960)

[20] VOLGIN, L. N. *The Fundamentals of the Theory of Controlling Machines.* (Monogr.) Sovetskoe Radio, 1962

Most Recent Development of Dynamic Programming Techniques and Their Application to Optimal Systems Design

R. L. STRATONOVICH

Introduction. Block-diagram of an Optimal Controller

As is known[1–4], dynamic programming theory solves, in principle, a large number of the problems connected with optimal systems synthesis. The applicability of dynamic programming methods is not impaired by taking into account white gaussian noise and other random factors in various components—the statistical nature of the signal to be reproduced, imprecise knowledge of it, random influences on the controlled plant, or interference in the feedback circuit (*Figure 1*). Of course, as the problems grow more complicated, the actual performance of the calculations becomes more and more difficult.

Although the basic principles of dynamic programming were expounded long ago, the number of non-trivial problems of optimal control theory actually solved by this method is not large. This is explained by purely computational difficulties which have to be overcome before a solution is found.

What has been said confirms the importance of the development of new methods and techniques to increase the effectiveness of the theory and make it easier for concrete results to be obtained.

In complex statistical problems the effective use of the theory becomes possible as a result of the introduction of 'sufficient coordinates' on which the risk function depends. The importance of this concept was noted by Bellman and Kalaba[2], and the author has clarified and developed it further[5, 6].

The sufficient coordinates form the space in which the Bellman equation is considered. A non-trivial statistical example is used in this paper to illustrate the effectiveness of the introduction of sufficient coordinates. In the example, the sufficient coordinates are a combination of *a posteriori* probabilities and the dynamic variables of the controlled plant.

In complex statistical problems the introduction of sufficient coordinates has the result that the optimal controller breaks down into at least two consecutive units, each of which is constructed according to its own principles. The first unit *SC* (*Figure 1*) produces the sufficient coordinates $\vec{X}$. In some dynamic programming problems it is trivial, but in complex statistical problems it may perhaps prove most important. In the latter, it is synthetized with the aid of methods similar to those of non-linear optimal filtration[7]. In the example considered below, it simply coincides with a unit effecting optimal non-linear filtration.

The signals from the *SC* unit output are sent to a further unit *OC*, which produces the optimal control action. The form of this unit, which converts the sufficient coordinates into a control signal, is found by consideration of the Bellman equation. This unit can be synthetized without great difficulty if the risk function is first found as a solution of the Bellman equation. The most difficult problem is the obtainment of this solution. Therefore, techniques and methods, which make it easier to obtain the solution of this equation, are of interest.

The equation is made far simpler by considering the stationary mode of operation, when the time-dependence and time-derivative are eliminated from the Bellman equation. The corresponding stationary equation was considered by Stratonovich and Shmalgauzen[8], and the method quoted is also described in this paper. Furthermore, to solve the resultant equation, use is made of the asymptotic step-by-step approximation method, first expounded by the author[9]. This method is convenient for the case of small diffusion terms, and makes it possible to obtain consecutive approximations whose accuracy is determined by the magnitude of the coefficients for the second derivatives in the Bellman equation.

It must be noted that the number of methods for approximate solution of the Bellman equation, which can be thought up for the solution of concrete problems, is practically unlimited; each method is best suited for the solution of problems of a particular type. To them must be added the obtaining of a solution on analogue or digital computers. Out of the whole range of methods, a special approximate method will be described and applied to the example under consideration, in the concluding part of the paper. The essence of this method is that the risk function is represented as a function whose appearance is fully determined by a finite number of parameters $\vec{a}$. The Bellman equation for the risk function is replaced by a system of equations which specify the evolution of these parameters in inverse time. This system is roughly equivalent to the original Bellman equation.

The unit *OP* (*Figure 1*) simulates this system of equations and determines the parameters $\vec{a}$ as a function of time. It operates as a self-contained unit, if measurement of the statistics of the processes and other variables is not carried out in the course of operation, and must finish its work before the start of operation of the main system. If the operating conditions change, then there may be a need for periodic plotting of the process of determination of the parameters by the *OP* unit in application to the new operating conditions. Such a system will belong to the class of adaptive systems. The *OC* unit produces the optimal control action in response to the values of the sufficient coordinates and the risk function parameters corresponding to a given moment of time. The corresponding algorithm is derived from the form of the Bellman equation and the adopted approximation of the risk function.

Usually the transition to a finite number of parameters entails some deterioration of the quality of operation of the system. The greater the number of parameter taken, the higher the accuracy

of approximation and the closer the system to optimal, but, on the other hand, the more complicated the *OP* unit. For a specified number of parameters is is important to determine the successful choice of the means of approximation. Here a great deal depends on the ingenuity and inventiveness of the designer. In this paper, one natural means of selecting the parameters is suggested—taken as the parameters are the bottom coefficients of the expansion of the risk function by a suitable full set of functions.

The block diagram of an optimal controller given in the paper is of a basic nature, and in fact not all the units need be there. In some problems the *SC* unit can be left out because of triviality. The *OP* unit can be separated from the system. It can be replaced by a preliminary calculation, and the parameter values can be taken into account once and for all in the synthesis of the *OC* unit. The situation is different if the system itself investigates varying conditions of operation. In that case to the units *OP*, *SC*, *OC* (if there is no *OP* unit) must be sent the signals from the appropriate metering devices.

Example—Sufficient Coordinates—Stationary Fluctuation Regime

Let the variable part of the system—the controlled plant *CP* (*Figure 1*)—have a transfer function $K(p)$. Let the control action u be limited to the values $-1 \leq u \leq 1$. The input signal x_t, like the output signal y_t, is assumed to be known accurately. Let the signal on the input $x_t = s_t + \xi_t$ be the sum of the pulse signal $s_t = \pm 1$ and interference be the normal white noise ξ_t ($M\xi_t = 0$; $M\xi_t \xi_{t+\tau} = \varkappa\delta(\tau)$).

The task of the system is to ensure that the coordinate of the plant y_t reproduces as accurately as possible the pulse signal s_t. If $s_t = 1$, but $y_t \neq 1$, the penalty $c(1, y_t)$ in a unit of time is taken. The functions $c(\pm 1, y_t)$ can differ. For the step-by-step method, which is used to obtain formula (22), the condition that these functions be differentiable is essential. Henceforward, to make things specific, use will be made of the criterion of the minimum mean square error, which corresponds to the functions

$$c(s, y) = (s-y)^2 \tag{1}$$

It will be assumed that the signal s_t is *a priori* a symmetrical two-position Markovian process, moreover the *a priori* probabilities $p_t(\pm 1) = P[s_t = \pm 1]$ satisfy the equations

$$\frac{dp(1)}{dt} = -\frac{dp(-1)}{dt} = -\mu p(1) + \mu p(-1) \tag{2}$$

This means that the pulses and intervals are independent and distributed according to the exponential law $P[\tau > c] = e^{-\mu c}$.

It is required to design an optimal controller which produces a control signal u_t so that the mean penalties are reduced to a minimum. The latter is a function of the sufficient coordinates.

The sufficient coordinates of the given problem will be considered. Their definition, which is given by the author[5, 6] reduces to the requirement of the sufficiency of the selected coordinates in three respects:

(*a*) Sufficiency for determination of the conditional mean penalties:

$$r_t = M[c_t | x_\tau, u_\tau, \tau < t] \tag{3}$$

(*b*) Sufficiency for indication of the constraints of choice of the control solution and (*c*) sufficiency for determination of the future evolution of the sufficient coordinates themselves (for the determination of the probabilities of their future values).

In the given problem the limitations of choice $|u_t| \leq 1$ at each moment of time t depend on nothing at all, so point (*b*) can be disregarded. Point (*a*) will be considered, and the *a posteriori* probabilities $w_t(\pm 1) = P[s_t = \pm 1 / x_\tau, \tau < t]$ introduced. Then the mean penalities (3) will be written

$$r_t = c(1, y)\, w_t(1) + c(-1, y)\, w_t(-1)$$

Requirement (*a*) will obviously be satisfied if the sufficient coordinates include the coordinate y and also the *a posteriori* probability or a magnitude replacing it, say $z = w(1) - w(-1)$.

The evolution of the variables of the given problem will be considered. The equation determining the behaviour of y_t depends on the appearance of the function $K(p)$. Obviously

$$\frac{dy_t}{dt} = u_t + n_t \text{ with } k = \frac{1}{p} \tag{4}$$

and

$$\frac{dy_t^2}{dt} + \rho \frac{dy_t}{dt} = \rho u_t + \rho n_t \tag{5}$$

with

$$k(p) = \frac{\rho}{p^2 + \rho p}$$

Assume that n_t is normal white noise ($Mn_t = 0$; $Mn_t n_{t+\tau} = N\delta(\tau)$). Then in case (4), y_t will be (with the fixation of $\{u_\tau\}$) a Markovian process, and the probability of the future values $y_{t+\Delta}$ will be entirely determined by the value at the present moment of time. In case (5), the two-dimensional process (y_t, dy_t/dt) is Markovian. The probability of the future values is determined by these two magnitudes y_t, dy_t/dt, and therefore the sufficient coordinates must necessarily include, apart from y_t, dy_t/dt for satisfaction of requirement (*c*). If, for example, the interference $\{n_t\}$ would be a unidimensional Markovian process, then n_t should be be included among the sufficient coordinates.

The mode of variation of z_t is now found, and it is proved that it does not require the introduction of new sufficient coordinates. The variation of the *a posteriori* probabilities is induced by two causes—*a priori* transfers between states $s = 1$, $s = -1$, and also variation of the *a posteriori* probabilities as a result of supplementary observation of the process x_t. If there were no observation, the probabilities $w_t(\pm 1)$ would vary in accordance with eqns (2):

$$\begin{aligned} \frac{dw_t(1)}{dt} &= -\mu w_t(1) + \mu w_t(-1) \\ \frac{dw_t(-1)}{dt} &= \mu w_t(1) - \mu w_t(-1) \end{aligned} \tag{6}$$

If there were no *a priori* transfers, the *a posteriori* probabilities after the observation $x_\tau = s + \xi_\tau$ in the interval $t_0 \leq \tau \leq t$ could be expressed through the probabilities $w_0(\pm 1)$, before this observation in accordance with the Beiss formula

$$w_t(s) = \text{const}\, w[\xi_\tau, t_0 \leq \tau \leq t]_{\xi_\tau = x_\tau - s} \cdot w_0(s) \tag{7}$$

Here $w[\xi_\tau]$ is the probability distribution for $\{\xi_\tau,\ t_0 \leq \tau \leq t\}$ which for white noise, as is known, has the form

$$w[\xi_\tau] = \text{const} \exp\left[-\frac{1}{2\kappa}\int_{t_0}^{t} \xi_\tau^2\, d\tau\right]$$

Substituting into this $\xi_\tau = x_\tau - s$, and relating $-1/2\kappa\int_{t_0}^{t}(x_\tau^2+1)\,dt$ to the multiplier C, which does not depend on y, in accordance with (7), gives

$$w_t(s) = C\exp\left[\frac{1}{\kappa}\int_{t_0}^{t} x_\tau s\, d\tau\right]\cdot w_0(s)$$

From this, differentiation according to t gives

$$\frac{dw_t(s)}{dt} = \left[\frac{1}{C}\frac{dC}{dt} + \frac{x_t s}{\kappa}\right] w_t(s) \tag{8}$$

Returning to the case of the *a priori* transfers, eqns (6) and (8) must be combined. This gives

$$\frac{dw(1)}{dt} = -\mu w(1) + \mu w(-1) + \left[\frac{x_t}{\kappa} + \frac{1}{C}\frac{dC}{dt}\right] w(1)$$

$$\frac{dw(-1)}{dt} = \mu w(1) - \mu w(-1) + \left[-\frac{x_t}{\kappa} + \frac{1}{C}\frac{dC}{dt}\right] w(-1) \tag{9}$$

The derivative $1/c\ dc/dt$ is determined from the condition of retention of the norm $d/dt\,[w_t(1) + w_t(-1)] = 0$ and proves equal to $-x_t/\kappa\,|\,[w(1) - w(-1)]$. Substituting this value into (9) and transferring to the variable $z = w(1) - w(-1)$, gives the equation

$$\frac{dz}{dt} = -2\mu z + \frac{1-z^2}{\kappa} x_t \tag{10}$$

which was derived by the author[9] on the basis of the general theory.

Since in (10) $x_t = s_t + \xi_t$, and ξ_t is white noise, the probabilities of the future values are determined by the value of z_t and the behaviour of s_t, $\tau > t$. But since s_τ is a Markovian process, its behaviour is determined by the value of s_t, which is described by the probabilities $w_t(s_t)$, that is to say, once again by the coordinate z_t. Hence the introduction of new variables in accordance with requirement (*c*) is not necessary.

Equations (4), (5) and (10) make it possible to write an alternative equation or Bellman equation for the given problem. Case (4) will be dealt with first. Introducing the function of minimum future risks

$$\boldsymbol{S}(y,z,t) = \min_{u_\tau,\ \tau \geq t} M\left\{\int_t^T \boldsymbol{C}_\tau\, d\tau \,\middle|\, y_t, z_t\right\} \tag{11}$$

(T is the time of termination of operation), and compiling the difference of these expressions for the two moments t and $t + \Delta$, gives the equation

$$\frac{\partial \boldsymbol{S}(y,z,t)}{\partial t} + \lim_{\Delta\to 0} \min_{u_\tau} M\left\{\frac{S(y_{t,\Delta} z_{t+\Delta}, t) - S(y_t, z_t, t)}{\Delta} + \boldsymbol{C}_t \,\middle|\, y_t, z_t\right\}$$
$$(t \leq \tau < t+\Delta) \tag{12}$$

In computing the limit which stands here, a Taylor expansion by the increments $y_{t+\Delta} - y_t$, $z_{t+\Delta} - z_t$ will be performed and both the linear and quadratic terms will be taken into account. The differentiability of the risk function is assumed. Eqn (4) gives

$$\lim_{\Delta\to 0} M\frac{y_{t+\Delta} - y_t}{\Delta} = u_t;\ \lim_{\Delta\to 0} M\frac{(y_{t+\Delta} - y_t)^2}{\Delta} = N \tag{13}$$

Computation of the Fokker-Planck coefficients for the second coordinate z_t is somewhat more complicated. In the process, the equality

$$M\{x_t | z_t\} = M\{s_t | z_t\} = z_t$$

must be taken into account, eqn (10) must be used, and the well-known technique of averaging stochastic eqns (10) must be applied. The result of the averaging has the form

$$\lim_{\Delta\to 0} M\left\{\frac{z_{t+\Delta} - z_t}{\Delta}\,\middle|\, z_t\right\} = -2\mu z_t$$
$$+\frac{1-z^2}{\kappa} M\{x_t | z_t\} + \frac{\partial}{\partial z_t}\left(\frac{1-z_t^2}{\kappa}\right)\frac{1-z_t^2}{z} = -2\mu z_t \tag{14}$$

Moreover

$$\lim_{\Delta\to 0} M\frac{1}{\Delta}(y_{t+\Delta} - y_t)(y_{t+\Delta} - y_t) = 0$$

$$\lim_{\Delta\to 0} M\left\{\frac{(z_{t+\Delta} - z_t)^2}{\Delta}\,\middle|\, z_t\right\} \quad \frac{(1-z_t^2)^2}{\kappa^2} \lim_{\Delta\to 0} M\frac{(x_{t+\Delta} - x_t)^2}{\Delta}$$
$$= \frac{(1-z_t^2)^2}{\kappa} \tag{15}$$

Hence, eqn (12) adopts the form

$$\frac{\partial S}{\partial t} + \min\left[\pm\frac{\partial S}{\partial y}\right] - 2\mu z\frac{\partial S}{\partial z} + \frac{N}{2}\frac{\partial^2 S}{\partial y^2} + \frac{(1-z^2)^2}{2\kappa}\frac{\partial^2 S}{\partial z^2}$$
$$+ C(1,y)\frac{1+z}{2} + C(-1,y)\frac{1-z}{2} = 0 \tag{16}$$

The second term can also be written in the form $-|\partial S/\partial y|$. To the resultant eqn (16) must be added the boundary conditions. In view of the fact that $|s| \leq 1$, only the domain $|y| \leq 1$ need be considered. Because (16) contains the diffusion term $1/2\,N\,\partial^2 S/\partial y^2$ on the boundaries $y = \pm 1$ there must hold the conditions

$$\frac{\partial S}{\partial y}(\pm 1, z, t) = 0 \tag{17}$$

Since $0 \leq w(s) \leq 1$, for the second coordinate one has $|z| \leq 1$. On the sides $z = \pm 1$ of the square the diffusion coefficient for the second diffusion number $1/2\,\kappa\,(1-z^2)^2\,\partial^2 S/\partial z^2$ vanishes. Therefore, instead of the conditions $\partial S/\partial z = 0$ on these sides the more trivial conditions

$$\left|\frac{\partial S}{\partial z}(y, \pm 1, t)\right| < \infty \tag{18}$$

are satisfied.

R_+ will be used to denote the domain of the space of the sufficient coordinates, where $\partial S/\partial y > 0$, and correspondingly R_- where $\partial S/\partial y < 0$. The boundary Γ between R_+ and R_- will be termed the switching line or separatrix; it is to the finding of this line that the calculation of the OC unit (*Figure 1*) reduces. On it are satisfied the conditions of continuity of the

risk function and its first derivatives $\partial S/\partial y$, $\partial S/\partial z$. These conditions are a consequence of the diffusion nature of eqn (16). From the continuity of the derivative $\partial S/\partial y$ there follows the condition

$$\frac{\partial S}{\partial y}=0 \text{ on } \Gamma \tag{19}$$

Eqn (16) describes the evolution of the risk function with the inverse passage of time. The role of the initial condition for it is played by the fixation of the risks at the moment of termination of the operation $S(y, z, T)$. If there are no special additional considerations, then $S(y, z, T)$ can be made equal to zero.

The Bellman equation is also derived in a similar way for more complex functions $K(p)$. As in case (5), the velocity $v=\partial y/\partial t$ must be included among the sufficient coordinates. Then the function $S(y,v,z,t)$ will satisfy the equation

$$\frac{\partial S}{\partial t}+v\frac{\partial S}{\partial y}-\rho v\frac{\partial S}{\partial v}-\rho\left|\frac{\partial S}{\partial v}\right|-2\mu z\frac{\partial S}{\partial z}+\frac{\rho_2 N}{2}\frac{\partial^2 S}{\partial v^2}$$
$$+\frac{(1+z^2)^2}{2\kappa}\frac{\partial^2 S}{\partial z^2}+C(1,y)\frac{1+z}{2}+C(-1,y)\frac{1+z}{2}=0 \tag{20}$$

An important particular problem among the group of problems connected with optimal systems synthesis is the problem of calculating the optimal stationary mode of operation. In this case the operation-termination time T tends to infinity. Then, irrespective the values of the coordinates at the moment t a stationary fluctuation mode is established in the system, characterized by some mean penalty γ in a unit of time. This means that when T increases, e.g., by Δt, the risk function increases by $\gamma\,\Delta t$

If the difference $S(t)-\gamma(T-t)$ is formed and the limit transfer $T\to\infty$ performed, the resultant function will not depend on time. In case (4) this function

$$f(zy)=\lim_{T\to\infty}[S(y,z,t)-\gamma(T-t)]$$

as can easily be seen in accordance with (16) satisfies the equation

$$\left|\frac{\partial f}{\partial y}\right|+2\mu z\frac{\partial f}{\partial z}=\frac{N}{2}\frac{\partial^2 f}{\partial y^2}+\frac{1}{2\kappa}(1-z^2)^2\frac{\partial^2 f}{\partial z^2}+y^2-2yz+1-\gamma \tag{21}$$

[here (1) is used]. Moreover the same conditions (17)–(19) are satisfied on the boundaries as before. The solution of eqn (21) makes it possible to find simultaneously the function $f(y,z)$, the switching line Γ and the stationary mean penalty γ. The same holds for eqn (20).

Solving the Bellman Equation

In view of the difficulty of obtaining a precise solution of the alternative equation, various approximate methods can be developed. Some of them will be illustrated, taking eqns (16) and (21) as an example. Of course the methods—for example, the method of parameters—permit generalization to other more complex cases as, say case (20), but then the laboriousness of the calculations increases markedly. The results obtained with the aid of (16) are also approximately valid for case (20), when $\varrho\gg 1$, i.e., when the inertia of the controlled plant plays a small part and can be disregarded.

In this case, the optimal control action depends on the variables y, z, and equals $u=1$ in the domain R_+ (correspondingly, $u=-1$ in R_-). *Figure 2* shows the approximate location of these domains, and of the switching line; the mean transfer velocities $M\,dy/dt$, $M\,dz/dt$ are also given. An approximate calculation was performed of the switching line in the stationary case (eqn 21), by the asymptotic step-by-step method developed by the author[9]. For the case $N=0$, $2\mu<1$, the switching line of the first approximation was found to be

$$z_t(y)=y+\frac{2\mu y}{\kappa}\frac{(1-y^2)^2}{1-4\mu^2y^2} \tag{22}$$

The higher approximations have an order of $(\mu/\kappa)^2$ and higher.

The second approximate method of solution, which has a wider sphere of application, will be dealt with in greater detail. This method is linked with the determination of the parameters of the risk function, to which corresponds the unit *OP* in *Figure 1*, as was stated in the introduction.

One of the ways of introducing the parameters is the expansion of the risk function according to some preselected suitable system of functions. For the given example these are the functions of the variables y and z. Let $\varphi_0(y)$, ..., $\varphi_{r-1}(y)$ and $\psi_0(z)$, ..., $\psi_{s-1}(z)$ be the selected functions. Then the parameters of the risk function will be the coefficients $a_{ij}(t)$ of the expansion

$$S(y,z,t)\sim\sum_{i=0}^{r-1}\sum_{j=0}^{S-1}a_{ij}(t)\,\varphi_i(y)\,\psi_j(z) \tag{23}$$

Since the above systems of functions are not complete, replacement of the risk function by the expression given usually entails some errors. To make the coefficients a_{ij} more exact, any criterion is set, e.g., the minimum integral from the square of the difference

$$\int_{-1}^{1}\int_{-1}^{1}\left[S-\sum_{ij}a_{ij}\varphi_i\psi_j\right]^2 dy\,dz=\min$$

will be required.

The variation of this expression leads to a system of linear equations

$$\sum_{i,j}a_{ij}(\varphi_i,\varphi_e)(\psi_j\psi_m)=(S,\varphi_e\psi_m) \tag{24}$$
$$e=0,\ldots,r-1;\ m=0,\ldots,s-1$$

which permits a_{ij} to be calculated, if $S(y,z,t)$ is known.

Here is written

$$(\varphi_i,\varphi_e)=\frac{1}{2}\int_{-1}^{1}\varphi_i\varphi_e\,dy;$$

$$(S,\varphi_e\psi_m)=\frac{1}{4}\int_{-1}^{1}\int_{-1}^{1}S\,\varphi_e\psi_m\,dy\,dz$$

With the aid of the inverse matrices

$$\|c_{ie}\|=\|(\varphi_i,\varphi_e\|^{-1};\ \|c'_{jm}\|=\|(\psi_j,\psi_m\|^{-1} \tag{25}$$

the solution of system (33) can be written as

$$a_{ij}=\sum_{e,m}c_{ie}\,c'_{jm}(S,\varphi_e\psi_m) \tag{26}$$

How the equation for the parameters is obtained from the alternative equation will now be shown. Let the latter have the form

$$\frac{\partial S}{\partial t}=\mathscr{F}[S] \tag{27}$$

Differentiating (26) according to time, and substituting (27) into the right-hand side gives

$$\frac{da_{ij}}{dt}=\sum_{e,m} c_{ie}c'_{jm}(\mathscr{F}[S],\varphi_e\psi_m)$$

If the replacement of (23) is performed here, this will give a closed system of equations for the parameters

$$\frac{da_{ij}}{dt}=\sum_{e,m} c_{ie}c'_{jm}(\mathscr{F}[\sum_{p,q} a_{pq}\varphi_p\psi_q],\varphi_e\psi_m) \tag{28}$$

The example being considered will be utilized to illustrate the application of this method. Because of the boundary condition (17) it is convenient to select the functions $\varphi_i(y)$, each of which possesses this property $d\varphi_i/dy\,(\pm 1)=0$. For the second coordinate z, there is no such condition, so

$$r=s=3;\ \varphi_0(y)=\psi_0(y)=1;\ \varphi_1(y)=\sqrt{2}\sin\frac{\pi y}{2};$$

$$\varphi_2(y)=\sqrt{2}\cos\pi y;\ \psi_1(z)=z;\ \psi_2(z)=z^2$$

can be written.

In the given case $(\varphi_i,\varphi_e)=c_{ie}=\delta_{ie}$;

$$\|(\psi_j,\psi_m)\|=\begin{Vmatrix}1 & 0 & \frac{1}{3}\\ 0 & \frac{1}{3} & 0\\ \frac{1}{3} & 0 & \frac{1}{5}\end{Vmatrix};\ \|c'_{jm}\|=\begin{Vmatrix}\frac{9}{4} & 0 & -\frac{15}{4}\\ 0 & 3 & 0\\ -\frac{15}{4} & 0 & \frac{45}{4}\end{Vmatrix} \tag{29}$$

Since the risk function is symmetrical $S(y,z,t)=S(-y,-z,t)$ (with symmetrical penalties $S(y,z,T)$, then in expansion (23) there should be present only symmetrical terms

$$S(y,z,t)\sim a_{00}+a_{02}z^2+a_{11}z\sqrt{2}\sin\frac{\pi}{2}y$$
$$+(a_{20}+a_{22}z^2)\sqrt{2}\cos\pi y \tag{30}$$

Moreover, putting $a_{20}=\alpha a_{11}$; $a_{11}=\beta a_{21}$, it is expedient to make the substitution

$$\left|\frac{\partial S}{\partial y}\right|=\frac{\pi}{\sqrt{2}}|a_{11}|\left|z\cos\frac{\pi}{2}y-2(\alpha+\beta z^2)\sin\pi y\right|$$
$$\sim\frac{\pi}{\sqrt{2}}|a_{11}|\sum_{ij}\rho_{ij}(\alpha,\beta)\varphi_i(y)z^j \tag{31}$$

where

$$\rho_{ij}(\alpha,\beta)=\sum_{e,m} c_{ie}c'_{jm}\sigma_{em} \tag{32}$$

$$\sigma_{em}=\frac{1}{4}\int_{-1}^{1}\int_{-1}^{1}\left|z\cos\frac{\pi}{2}y-2(\alpha-z^2\beta)\sin\pi y\right|\varphi_e(y)z^m\,dy\,dz$$

In addition, within the framework of the selected approximation

$$(1-z^2)^2\sim\frac{32}{35}-\frac{8}{7}z^2;\ y\sim\frac{8}{\pi^2}\sin\frac{\pi}{2}y$$

$$y^2\sim\frac{1}{3}-\frac{4}{\pi^2}\cos\pi y$$

After the above substitutions, eqn (16), where $\frac{1}{2}c(1,y)(1+z)+\frac{1}{2}c(-1,y)(1-z)=y^2-2yz+1$ adopts the form

$$\sum_{i,j}\frac{da_{ij}}{dt}\varphi_i z^j=\frac{\pi}{\sqrt{2}}|a_{11}|\sum_{ij}\rho_{ij}\varphi_i z^j$$
$$+2\mu[2a_{a2}z^2+a_{11}z\varphi_1+2a_{22}z^2\varphi_2]$$
$$-\frac{1}{\kappa}\left(\frac{32}{35}-\frac{8}{7}z^2\right)[a_{02}+a_{22}\varphi_2]$$
$$+\frac{N}{2}\frac{\pi^2}{4}[a_{11}z\varphi_1+4(a_{20}+a_{22}z^2)\varphi_2]$$
$$-\frac{1}{3}+\frac{2\sqrt{2}}{\pi^2}\varphi_2+\frac{8\sqrt{2}}{\pi^2}z\varphi_1-1$$

Separately equating the coefficients of the functions $\varphi_i z^j$ gives five equations for the da_{ij}/dt derivatives. The most important of these are the three equations

$$\frac{da_{11}}{dt}=\frac{\pi}{\sqrt{2}}|a_{11}|\rho_{11}\left(\frac{a_{20}}{a_{11}},\frac{a_{22}}{a_{11}}\right)+2\mu a_{11}+\frac{\pi^2}{8}Na_{11}+\frac{8\sqrt{2}}{\pi^2}$$

$$\frac{da_{20}}{dt}=\frac{\pi}{\sqrt{2}}|a_{11}|\rho_{20}\left(\frac{a_{20}}{a_{11}},\frac{a_{22}}{a_{11}}\right)-\frac{32}{35}\frac{a_{22}}{\kappa}+\frac{\pi^2}{2}Na_{20}+\frac{2\sqrt{2}}{\pi^2}$$

$$\frac{da_{22}}{dt}=\frac{\pi}{\sqrt{2}}|a_{11}|\rho_{22}\left(\frac{a_{20}}{a_{11}},\frac{a_{22}}{a_{11}}\right)+4\mu a_{22}+\frac{8}{7}\frac{a_{22}}{\kappa}+\frac{\pi^2}{2}Na_{22} \tag{33}$$

The switching line is found by equating to zero the derivative (31). The equation of this line has the form

$$4(\alpha+\beta z_\Gamma^2)\sin\frac{\pi}{2}y=z_\Gamma;\ z_\Gamma=z_\Gamma(y) \tag{34}$$

The course of the switching line is determined only by the relations $\alpha=\frac{a_{20}}{a_{11}}$, $\beta=\frac{a_{22}}{a_{11}}$ of the parameters entering into (33).

As is usual in dynamic programming, eqn (33) must be solved for the inverse passage of time. If the inverse time $t_1=T-t$ is introduced, the conditions corresponding to the end of operation will look like 'initial' conditions. In the absence of conclusive penalties at the moment T the corresponding conditions will be null:

$$a_{11}=a_{20}=a_{22}=0 \text{ when } t_1=0\left(\alpha=\frac{1}{4},\beta=0\right)$$

When a sufficiently long time t passes, the mode of operation of the system approaches the stationary. This corresponds to the approach of the parameters a_{11}, a_{20}, a_{22} to the stationary values a_{11}^0, a_{20}^0, a_{22}^0. The latter are the solution of the system of three equations obtained by equating to zero expressions (33).

Using (29) and (34), formulae (32) can be brought to the form

$$\rho_{i1}(\alpha,\beta)=3\sigma_{i1};$$

$$\rho_{10}(\alpha,\beta)=\frac{9}{4}\sigma_{i0}-\frac{15}{4}\sigma_{i2};$$

$$\rho_{i2}(\alpha,\beta)=\frac{45}{4}\sigma_{i2}-\frac{15}{4}\sigma_{i0};$$

$$\sigma_{ij}=\frac{1}{2}\int_{-1}^{1}\left[\frac{1-z_r^{j+2}}{j+2}\cos\frac{\pi}{2}y-2\left(\alpha\frac{1-z_r^{j+1}}{j+1}\right.\right.$$

$$\left.\left.+\beta\frac{1-z_r^{j+3}}{j+3}\right)\sin\pi y\right]\varphi_i(y)\,dy \qquad (35)$$

For further calculation of the functions $\varrho_{ij}(\alpha,\beta)$ numerical methods can be employed, or use can be made of one or another approximation of the function $z_r(y)$.

The solution of the given problem consists in the fact that the unit *OP* (*Figure 1*) realizes eqns (33) in inverse time, and unit *OC* realizes the switching line (34).

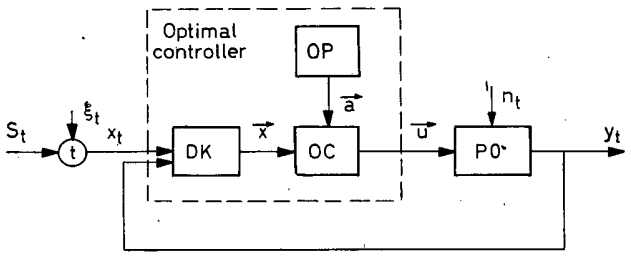


Figure 1. Optimal servosystem. SC: sufficient-coordinates unit; OP: parameter-determination unit; OC: optimal control unit; CP: controlled plant; DK = SC; 0⌐ = OP; OY = OC; PO = CP

References

[1] BELLMAN, R. *Dynamic Programming*. 1960. Moscow; Foreign Languages Publishing House

[2] BELMAN, R., and KALABA, R. Dynamic programming and feedback control. *Automatic and Remote Control*. Vol. 1, p. 460. 1961. London; Butterworths

[3] FELDBAUM, A. A. The theory of dual control, I–IV. *Automat. Telemech.* 21 (1960), 9, 11; 22 (1961), 12

[4] STRATONOVICH, R. L. Conditional Markovian processes in prolems of mathematical statistics and dynamic programming. *Dokl. Akad. Nauk. SSSR* 140 (1961)

[5] STRATONOVICH, R. L. On optimal control theory. Sufficient coordinates. *Automat. Telemech.* 23 (1962), 7

[6] STRATONOVICH, R. L. Conditional Markovian processes in problems of mathematical statistics, dynamic programming and games theory. *4th All-Union Math. Congr., Leningrad* (1961)

[7] STRATONOVICH, R. L. Conditional Markovian processes. *Probability Theory and Its Application*. 5 (1960)

[8] STRATONOVICH, R. L., and SHMALGAUZEN, V. I. Some stationary problems of dynamic programming. *Izv. Akad. Nauk SSSR, Otdel. Tekh. Nauk. Energ. Automat.* 5 (1962)

[9] STRATONOVICH, R. L. On the optimal control theory. An asymptotic method of solving the diffusion alternative equation. *Automat. Telemech.* 23 (1962), 2

[10] STRATONOVICH, R. L. Selected problems of the theory of fluctuations in radio engineering. *Sov. Radio* (1961)

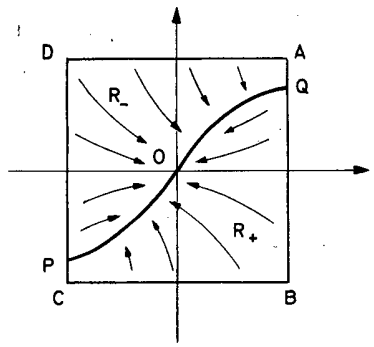


Figure 2. Space of sufficient coordinates. POQ: separatrix ⌐; POQBC: domain R_+; POQAD: domain R_-

The Realization of Optimal Programmes in Control Systems

G.S. POSPELOV

Methods of mathematical programming [the term is used to mean the application of mathematics to the practical activity of planning, development, decision-making etc., and is a natural generalization of such concepts as linear (or non-linear) dynamic programming] are spreading to all branches of the national economy, economics, engineering, industry, agriculture and so on. This presupposes the development of mathematical models of the events or sets of controlled plants which require to be controlled. Once the aim of control has been formulated, the task is to determine the optimum strategy of control whereby a programme of effects upon the controlled plants produces in some sense the optimal result.

It must be emphasized that the programming methods determine the strategy of control or *a priori* programme. The degree of coincidence between the actual result or process produced by control and the result or process anticipated from the *a priori* programme, is indicative, in particular, of the perfection of the mathematical model or of our knowledge about the controlled plant.

However, a mathematical model is a model and not the phenomenon itself, and, apart from this, during the process of realizing the *a priori* programme, the controlled plant can be affected by a variety of factors and perturbations which are not taken into account in the model. This can lead to deviations, and sometimes to substantial deviations, from the programme results, which by definition are optimal.

If the programme is time scheduled, use can be made of feedback to correct the effect of perturbations and inaccuracies in the mathematical description so as to ensure an actual programme closer to the optimal one.

The most completely represented by mathematical models are control systems. Taking their case as an example, we will consider the possible ways of realizing optimal programmes; in this instance, controll programmes.

A mathematical model of a control system is usually formed by means of ordinary differential equations. The control programme is broadly defined to cover the planning of the dynamic characteristics of the control system, its programme of operation, and the variation of the relationship. In all cases it is assumed that the system is provided with complementary feedbacks which improve the realization of the predetermined programme or *a priori* programme.

(1) The desired dynamic characteristics of a system are realized by complementary self-adjusting circuits, which in this case are complementary feedbacks which improve the realization of the predetermined programme of the control system of operation. *Figure 1* shows the well-known self-adjusting system of an automatic pilot which controls the angle of pitch of an aircraft[1]. The self-adjustment circuit changes the gain of the angular velocity circuit such that the margin of stability of this circuit is maintained constant. The correcting circuit 2 is selected to obtain a sufficiently high gain K. Under these conditions the transfer function of the closed angular velocity circuit is close to unity. Therefore, despite the variation of the properties of the controlled plant (owing to changes in flying conditions), the dynamic properties of the angle of pitch circuit will be determined by the transfer function of the model, i.e. in all cases they will be quite close to the predetermined or planned properties. Another example is the self-adjusting control system with extremal tuning of the correcting circuits[2]. Both examples refer to continuously operating control systems.

A somewhat special problem arises in the preservation of planned dynamic properties for 'single-action' systems[3] for which the behaviour is significant on a finite interval $t\,(0 \leq t \leq t_1)$, and for which the operating process is, as a rule, a transient process. Here one meets with the problem of maintaining a desired nature of transient behaviour, or a programme of motion of the representative point in phase space, on condition that the mathematical model does not exactly describe the dynamic properties of the controlled plant, nor the perturbations acting on the latter during the motion. Several possible ways of solving this problem are now indicated with simple examples.

Let the mathematical model of the controlled plant be represented in the form

$$\dot{x} = u \tag{1}$$

where x is the output coordinate and u is the controlling action.

Given the equation of the controller under the form

$$u = -a_0 x \tag{2}$$

the equation of the mathematical model of the system as a whole is

$$\dot{x} + a_0 x = 0 \tag{3}$$

Accordingly, for any initial condition x_0, the process of motion is characterized by an exponential with the exponent $-a$. Now suppose that there is a suspicion that, in fact, the control object is described by the equation

$$\dot{x} = f(x, u, t) \tag{4}$$

where

$$f(x, u, t) = -a(t) \cdot \phi(x) + F(t) + u \tag{5}$$

$a(t), u, F(t)$ are random functions and $\phi(x)$ is also random. Here it is known beforehand that $|\, a(t)\, \phi(x) + F(t)| < |\, u\,|$.

In this situation one can make a decision concerning the discrete control of the plant, such that at each step it is possible to control the fulfilment of the *a priori* programme, which is expressed as a function of time in the following manner:

$$x = x_0 e^{-at} \tag{6}$$

With discrete control we require for control a relationship between the value of $x(t)$ and the value of this coordinate at the instant of time $t + \Delta t$, i.e. the quantity $x(t + \Delta t)$. According to (6) this programme relationship is given by the relation

$$x(t+\Delta t) = x(t) \cdot e^{-a\Delta t} \tag{7}$$

where Δt is the interral of discreteness or the step of control. Using an analogy between the numerical solution of differential equations by difference methods and the discrete control of controlled plants, one writes the equation (5) in discrete form

$$x(t+\Delta t) = x(t) + f\left(t+\frac{\Delta t}{2}\right)\cdot\Delta t \tag{8}$$

where

$$f\left(t+\frac{\Delta t}{2}\right) = f\left[x\left(t+\frac{\Delta t}{2}\right), u\left(t+\frac{\Delta t}{2}\right), t+\frac{\Delta t}{2}\right]$$

The discrete form (8) of the solution of eqn (5) is used in the method proposed by Bashkirov. (The method of Bashkirov is described in the monograph by Popov[4].) According to eqn (8), by measuring the value of $x(t)$ at each step one can select the increment Δu at the instant $t + (\Delta t/2)$ such that $x(t + \Delta t)$ is governed by condition (7). The discrete form (8) is convenient in that the interval $\Delta t/2$ is available in the procedure for calculating $\Delta u(t + \Delta t/2)$. The information for calculating $\Delta u(t + \Delta t/2)$, apart from the known value of the desired $x(t + \Delta t)$, is obtained from the preceding values of Δu and x. In the general case $\Delta u(t + \Delta t/2)$ is calculated by the formula:

$$\Delta u\left(t+\frac{\Delta t}{2}\right) = \Delta u\left(t-\frac{\Delta t}{2}\right)\cdot\psi\left[x(t+\Delta t), x\left(t-\frac{\Delta t}{2}\right), x(t-\Delta t)\right] \tag{9}$$

The form of the function ψ depends on the particular theory of extrapolation which is adopted.

The information about the preceding values of x and u also includes information about changes in the properties of the object and of the perturbation $F(t)$. The use of this information for calculating $\Delta u(t + \Delta t/2)$ represents the additional feedback signals, or self-adjusting signals, and makes it possible to realize more accurately the desired programme of motion more exactly[6].

Equation (5) and its results can be generalized without difficulty to multi-dimensional systems of any order. In this case the equation of the controlled plant in the vector form is

$$\frac{dX}{dt} = f(X, U; t) \tag{10}$$

where X is the vector with the components x_i $(i = 1, 2, \ldots, n)$, f is the vector with the components f_i $(i = 1, 2, \ldots, n)$, and U is the control vector with the components u_i $(i = 1, 2, \ldots, \gamma)$; $\gamma \leq n$.

The maintenance of planned dynamic properties of single-action systems can also be realized by a continuous control. Suppose, for example, that the mathematical model of the controlled plant is written in the form

$$\ddot{x} + a_1\dot{x} = u \tag{11}$$

and $|u| \leq u_0$.

Suppose also that it is required to realize the system with maximum operating speed. According to Pontryagin's principle[5] of the maximum, the equation of the controller is of the form

$$u = -u_m \operatorname{sign}[x + f(a_1, \dot{x})] \tag{12}$$

However, there is a suspicion that in fact the controlled plant can be described by the equation

$$\ddot{x} + a_1^*(t)\cdot\dot{x} + a_0^*(t)\cdot x = u + F(t) \tag{13}$$

In view of the incomplete information about $a_1^*(t)$, $a_0^*(t)$ and $F(t)$ it is impossible to prescribe the control law of type (12) which ensure the maximum operating speed.

In view of this one proceeds as follows, forming the acceleration control circuit $\ddot{x} = n$ by means of the controlling action u (*Figure 2*). If the pass band of this circuit is sufficiently high the error $\varepsilon_n = n_{pr} - n$ will be close to zero and the programme acceleration will be equal to the actual acceleration. In more complex cases the acceleration control circuit, like the pitch angle control circuit (*Figure 1*), can be a self-adjusting circuit. If now the programme acceleration is close to the actual acceleration, any desired variation of the coordinate x and its derivative may be required. Thus, to form the system of maximum operating speed in accordance with the mathematical model (11), it is sufficient to put

$$\ddot{x}_{pr} = \dot{x} = -a_1\dot{x} - u_1 \operatorname{sign}[x + f(a_1, \dot{x})] \tag{14}$$

The block diagram which realizes (14) is shown in *Figure 3*. In expression (14) u_1 is always less than u_0 since some part of the control resource $u_0 = u_1$ goes to compensate the perturbation $F(t)$ and to compensate the difference between the coefficients $a_1^*(t)$ and $a_0^*(t)$ on the one hand and the coefficients of the mathematical model a_1 and $a_0 = 0$ on the other. Thus, at the expense of some reduction of operating speed (since $u_1 < u_0$) a definite realization of the programme for the optimum transient process is obtained.

Any other law of variation of the coordinate x can be required in this example. It may, for example, be required that the transient process should take place in accordance with the solution of a linear equation with constant coefficients

$$\ddot{x} + a_1\dot{x} + a_0 x = 0 \tag{15}$$

For this, it is obviously necessary to put $\ddot{x}_{pr} = -a_1\dot{x} - a_0 x$.

Figure 4 shows oscillograms which have been obtained on the electronic simulator for the case when $|a_0^*(t)| \leq 0{\cdot}05$; $a_1^*(t) \leq 1{\cdot}0$; $a_1 = 0{\cdot}4$; $a_0 = 0{\cdot}04$. The gain of the servo motor of the acceleration control circuit was taken as 10 1/sec. It will be seen from the oscillogram that the perturbation $F(t)$ and the fluctuations of the coefficients $a_1^*(t)$ and $a_0^*(t)$ have no effect on the course of the coordinate x which is governed by the solution of eqn (15).

The results explained by this example are also capable of very wide generalization. The generalization consists in that for a known indeterminacy of the properties of the controlled plant and of the acting perturbations it is advisable to organise a self-adjusting subsystem of rapidly varying coordinates of the controlled plant or of its higher-order derivatives. After the programme variation of the rapidly varying coordinates or of

their higher-order derivatives has been largely determined by this subsystem, the law governing the variation of the slowly varying coordinates or lower-order derivatives of the output quantity of the controlled plant can be built as desired. The additional feedbacks which make it possible to realize the required programme of dynamic properties of the system in the example under consideration are the feedbacks amongst which are the self-adjusting circuits for acceleration control.

Very often the realization of desired dynamic properties for single-acting systems is handicapped by unfavourable combinations of initial conditions. In non-linear systems these unfavourable combinations of initial conditions can lead to instability of the process for a given realization. The effect of unfavourable combinations of initial conditions can be eliminated by changing the initial values of the coordinates and by the formation of special signals which act on the system and which are functions of the initial conditions. Briefly, this means creating special feedbacks with respect to the initial conditions. The idea of using feedback with respect to the initial conditions has already been published in a paper by the author[6].

(2) In developing systems with programme control of the output coordinates of the controlled plant use may, to a large extent, be made of the foregoing ideas and methods which relate to the realization of programmed dynamic properties of control systems.

Suppose, for example, that it is required to vary according to the programme $g_{pr}(t)$ the output coordinate $x(t)$ of the controlled plant (*Figure 5*). For this the input of a closed system consisting of the controlled plant and the controller receives the programme signal $g_{pr}(t)$. For a system with a high pass band, if no perturbations are present, it is well known that $x \cong g_{pr}(t)$. However, a random perturbation which is not taken into account can considerably distort the desired programmed variation of $g_{pr}(t)$. In order to fulfil more accurately the programme, an additional feedback is formed (shown by the dotted line in *Figure 5*) and the programme correction circuit *abcdega* is thereby formed. The programme signal $g_{pr}(t)$ is compared with the actual signal and the difference signal acts at the input to the fundamental system *via* a self-adjusting correction circuit with a high gain W_K. The correction circuit may consist of the elements 2, *K*, 6, 7, 8, 9, 10 and 11 which are shown in *Figure 1*. Assuming, for the sake of simplicity, $W_k = K$, the following operator relationship is obtained between the input and output for the circuit of *Figure 5*:

$$x=\frac{\phi_{(p)}(1+K)}{1+K\phi(p)}g_{pr}(t)+\frac{\phi_f(p)}{1+K\phi(p)}F(t) \qquad (16)$$

where

$$\phi(p)=\frac{W_pW_0}{1+W_pW_0} \text{ and } \phi_f(p)=\frac{W_0}{1+W_pW_0}$$

and expression (16) can be written as

$$x=\frac{\phi(p)\left(\frac{1}{K}+1\right)}{\frac{1}{K}+\phi(p)}g_{pr}(t)+\frac{\phi_f(p)}{\frac{1}{k}+\phi(p)}\frac{1}{K}F(t) \qquad (17)$$

It will be seen from (17) that if $K \to \infty$

$$x=g_{pr}(t) \qquad (18)$$

independently of the action of the perturbation $F(t)$ and the fluctuations of the parameters of the controlled plant. It is understood that in this case condition (18) is fulfilled approximately since $K = \infty$ is not realizable in actual conditions.

Another example of programme control is the method of stabilizing acceleration (*Figure 2* and *3*) with subsequent construction of the desired programmed variation of the coordinate x_0 by means of a computer.

Using this method the 'logarithmic navigation'[7] can be realized when the acceleration according to the programme $k\dot{x}/x$, and consequently, the coordinate x is the solution of the differential equation

$$x\ddot{x}-k\dot{x}=0$$

A very important case of programme control is that when it is important to maintain a functional relationship between one coordinate and another. For example, the optimum programme, as regards operating speed, for the altitude and speed of an aircraft, as calculated, for instance, by the method of dynamic programming, is a programme in the coordinates H and V, i.e. it is given as a functional relationship $H_{pr} = H_{pr}(V_{pr})$ (*Figure 6*), both the quantities H and V here being the output coordinates of an aircraft controlled by the altitude rudder (the thrust of the engine is usually maximum in this case). The relationship $H_{pr} = H_{pr}(V_{pr})$ can always be represented parametrically:

$$H_{pr}=H_{pr}(t)$$

$$V_{pr}=V_{pr}(t)$$

The altitude control circuit H can now be formed by the usual method (*Figure 7*). If the system is unaffected by perturbations and the calculated characteristics of the aircraft coincide with the actual characteristics, and if the atmosphere through which the aircraft is flying remains standard, the completion of the programme $H_{pr}(t)$ will at the same time imply the completion of the programme $V_{pr}(t)$, and consequently of the programme relationship $H_{pr} = H_{pr}(V_{pr})$. However, if all the stated conditions are not fulfilled, the completion of $H_{pr}(t)$ will not generally imply the fulfilment of $V_{pr}(t)$, and consequently the completion of $H_{pr} = H_{pr}(V)$. For the planned programme $H_{pr} = H_{pr}(V)$ to be fulfilled with acceptable accuracy, it is necessary to introduce a programme correction circuit[8]. For this purpose the programme value of speed is compared with the actual speed and the difference in terms of the transfer function W_k changes the rate at which the programme is delivered, i.e. the speed of the clocks of the programme mechanisms H_{pr} and V_{pr} (*Figure 8*). As a result the speed of the clock mechanism of the programme is not uniform and the programmes H_{pr} and V_{pr} become functions of some irregularly varying argument τ, i.e. $H_{pr}(\tau)$ and $V_{pr}(\tau)$. Elimination of the argument τ again brings us back to the original relationship $H_{pr}(V_{pr})$. However, insofar as the rate of delivery of the programme signal H_{pr} at the input of the system conforms to the fulfilment of the speed programme, the accuracy of the realization of $H_{pr} = H_{pr}(V_{pr})$ is substantially increased. A similar circuit can be constructed for the motion of some controlled plant along a prescribed unperturbed trajectory $y_e = y_e(x_e)$ in the coordinates x, y (*Figure 9*). However, this report is confined to the plane problem. Suppose that the speed of the object is V and that the orientation of the speed vector

is characterized by the angle ψ. The obvious relationship between the coordinates x, y and the speed is expressed as follows

$$\dot{y} = V \sin\psi + W_y \tag{19}$$

$$\dot{x} = V \cos\psi + W_x \tag{20}$$

where W_y and W_x are perturbations in the form of speeds of displacement of the environment relative to the system of coordinates x, y. [In the formulae (19) and (20) the actual values of the coordinates of the controlled plant are used. The values of the desired unperturbed trajectory are denoted as x_e and y_e.] Consider the kinematic problem, i.e. suppose that the angle ψ of the speed vector can be arranged arbitrarily. On this assumption the control circuit for the coordinate y is formed. Here it is required that

$$\sin\psi = k_\varepsilon \varepsilon \tag{21}$$

where

$$\varepsilon = y_{pr} = y \tag{22}$$

the term $y_{pr} = y_{pr}(t)$ here being the programme value of the coordinate y, which does not coincide, as will be seen below, with the unperturbed value $y_e = y_e(t)$.

The equation for the coordinate y is found from the equations (19), (21) and (22):

$$\dot{y} + Vk_\varepsilon y = Vk_\varepsilon y_{pr}(t) + W_y \tag{23}$$

Assuming $y = y_e + \Delta y$, we obtain now the equation for the deviation Δy from the unperturbed motion

$$\Delta\dot{y} + Vk_\varepsilon \Delta y = Vk_\varepsilon (y_{pr} - y_e) - \dot{y}_e + W_y \tag{24}$$

It is worthwhile selecting the programme signal y_{pr} in accordance with the formula

$$y_{pr} = y_e + \frac{\dot{y}_e}{Vk_\varepsilon} \tag{24a}$$

For this value of the programme signal, eqn (24) becomes

$$\Delta\dot{y} + Vk_\varepsilon \Delta y = W_y \tag{25}$$

This implies that in the absence of the action W_y the deviation from the unperturbed projectory will tend to zero. A constant action will cause a constant error.

The control block diagram for the coordinate y is shown in *Figure 10*. It is obvious that a single control circuit according to the coordinate y cannot ensure the necessary control of the coordinate x or the fulfilment of the required programme $y_e = y_e(x_e)$. According to the circuit shown in *Figure 10*, the coordinate x varies according to the expression

$$x = V\int_0^t \cos\psi_e \,dt - V\int_0^t \tan\psi_e \cdot \Delta\dot{y}\,dt + \int_0^t W_x\,dt \tag{26}$$

The first term in eqn (26) is the desired unperturbed value of $x = x_e$, the second term can be limited, since it is determined by the error in the circuit for the stabilization of y, and the third term for $W_x = \text{const}$ will continuously increase. In order to realize the programme of motion along the unperturbed trajectory it is necessary to proceed in the same way as in the previous case (see *Figure 8*), i.e. it is necessary to form, by measuring the error $x_{pr} - x$, a signal which acts on the speed of the programme mechanism $y_{pr}(\tau)$ and $x_{pr}(\tau)$.

It should be noted that it is much simpler to correct the programme by varying the speed of the programme clocks if in the first example $dV/dt > 0$, and in the second example if $dx/dt > 0$. Generalizing, this method of correction to the programme of a system with n coordinates and γ controlling devices, we shall note that in this case the argument of control (the non-decreasing coordinate V in the first example, and the non-decreasing coordinate x in the second) should be any constant sign form of system derivative[9].

Frequently this form of coordinate originates naturally from the statement of the problem. For example, this is the case if it is required to control the ingredients of a mixture as a function of the volume of this mixture when this volume is varying in a monotonous way.

References

[1] MELLEN, D. Application of adaptive flight control. *Symposium on Self-adjusting Systems*. Rome, April 1962

[2] KRASOVSKII, A. A. The dynamics of continuous control systems with extremal self-adjustment of correcting devices. *Automatic and Remote Control*. 1961. London; Butterworths

[3] POSPELOV, G. S. Concerning the principles of construction of various types of self-adjusting control system. *Symposium on Self-adjusting Systems*, Rome. April 1962

[4] POPOV, YE. P. *Dynamics of Control Systems*. 1954. Moscow; GITTL

[5] PONTRYAGIN, L. S., BOLTYANSKII, V. G., and GAMKRELIDZE, R. V. *Mathematical Theory of Optimal Processes*. 1961. Moscow; Fizmatgiz

[6] POSPELOV, G. S. Various methods of improving the quality of processes of regulation and control. *Symposium on Use of Computer Engineering in Automation of Production* (in Russian), 1961. Moscow; Mashgiz

[7] Green, Logarithmic navigation for precise guidance of space vehicles. *IRE Trans.*, ANF-8, No. 2 (1961)

[8] KOZIOROV, L. M., and KOROBKOV, M. N. A method of stabilizing the functional relationship between two interrelated variables by means of one control device. *Iz. Akad. nauk S.S.S.R., OTN, Energetika i Avtomatika*, No. 4 (1961)

[9] LETOV, A. M. *The Stability of Non-linear Control Systems*. 1955. Moscow; GITTL

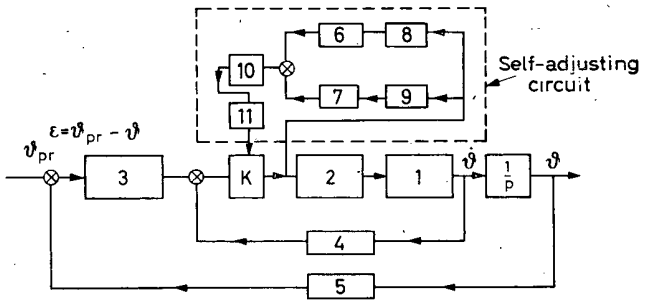


Figure 1. ϑ – angle of pitch; ϑ – programme value of pitch angle; 1 – object: 2 – correcting circuit; 3 – model; 4, 5 – measuring devices for the angular velocity $\dot{\vartheta}$ and the angle ϑ; 6, 7 – detectors; 8, 9 – high and low pass filters; 10 – servo motor; 11 – limiter

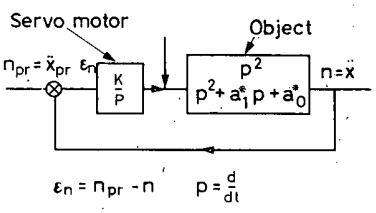


Figure 2

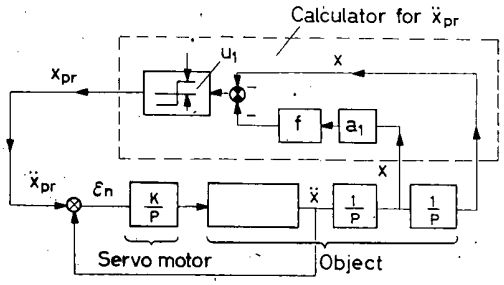


Figure 3

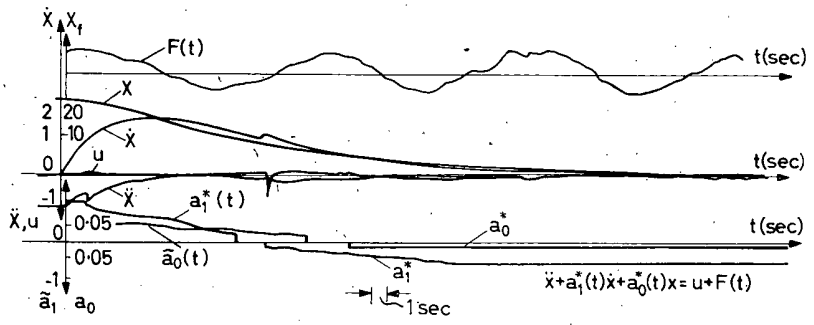


Figure 4

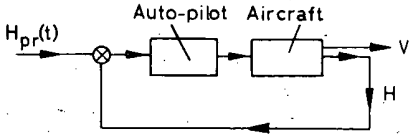


Figure 7

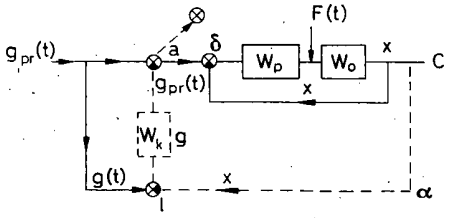


Figure 5

Figure 5. W_p – regulator; W_0 – object; W_k – self-adjusting correction circuit with a high amplification factor

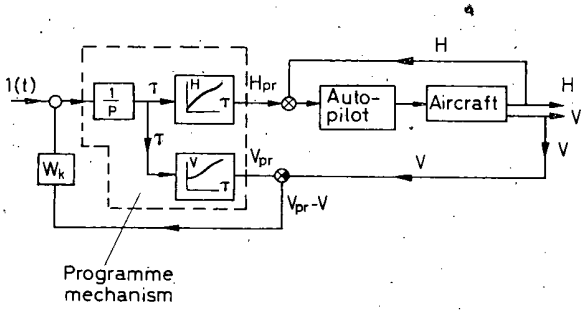


Figure 8

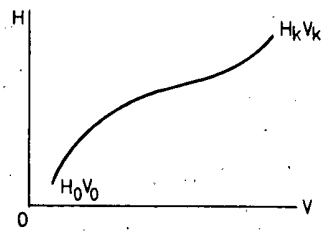


Figure 6

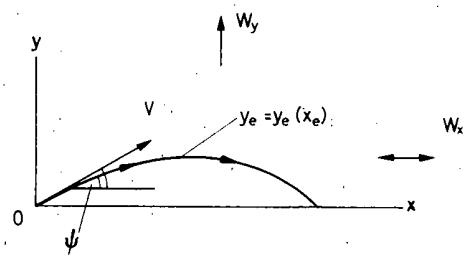


Figure 9