IFAC

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

WARSZAWA 1969

Control of Large Scale Systems

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

TECHNICAL SESSION
28

Organized by
Naczelna Organizacja Techniczna w Polsce

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

Control of Large Scale Systems

TECHNICAL SESSION No 28

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

Organized by
Naczelna Organizacja Techniczna w Polsce

Contents

Paper No			Page
28.1	USA	- M.D.Mesarovic, D.Macko, Y.Takahara - Two Coordination Principles and Their Application in Large Scale Systems Control.................	3
28.2	PL	- R.Kulikowski - Decentralized Optimalization of Large-Scale, Dynamic Systems................	22
28.3	SU	- A.A.Pervozvansky - Decentralization Principle at Optimalization of Complicated Systems.......	34
28.4	PL	- A.Straszak - On the Synthesis of Multi Level Large Scale Control Systems................	48
28.5	SU	- A.I.Kukhtienko - To Theory of Complicated Systems Controling............................	60

Wydawnictwa Czasopism Technicznych NOT
Warszawa, ul. Czackiego 3/5 — Polska

Zakład Poligraficzny WCT NOT. Zam. 34/69.

TWO COORDINATION PRINCIPLES AND THEIR APPLICATION IN LARGE SCALE SYSTEMS CONTROL

M.D. Mesarovic, D. Macko, Y. Takahara
Systems Research Center - Case Western Reserve University
Cleveland, Ohio USA

Two principles of coordination are formulated in order to provide guidance in selecting a structure for multi-level (hierarchical) control systems. One principle is based on interaction prediction and the other on interaction balance. Both are given within the general systems theoretic framework to emphasize their wide range of applicability. Sufficient conditions for the successful application of the principles are given for two-level systems defined on normed linear spaces. Some examples of two-level systems are given to illustrate the required conditions.

Introduction

Multi-level (hierarchical) systems are of increasing importance in the application of automatic control and automation. Indeed, the notion of a system with multi-level structure has become almost synonymous with the notion of a large scale system. A multi-level structure for a large scale system appears rather naturally in practice. It results from efforts toward the most efficient utilization of the available resources or inherent limitations of the elements out of which the system is being built.

Multi-level systems are being built in practice without any theoretical support on either a conceptual level in terms of the preferred structure or a detailed level in terms of synthesis and computational methods. To alleviate this situation a general theory of multi-level systems is being developed at the Systems Research Center of Case Western Reserve University. The result of this research is reported comprehensively in a forthcoming book. Presented here are some results on the synthesis of multi-level systems.

Consistency and Coordination

A multi-level control system consists of a class of control subsystems (units) arranged in a hierarchical fashion such that some of them have only indirect access to the controlled process. These control units receive information and manipulate or control other control units in the hierarchy. A two-level system with a single control unit on the second level is of special importance in the family of multi-level systems. It is the simplest system in the multi-level family and provides a convenient framework for the study of problems, such as the relationship between levels, which are characteristic of multi-level systems. In a two-level system the second level control unit is referred to as *supremal* while the first level control units are called *infimal*. The infimals are the only units which have direct access to the controlled process. More complex systems with many levels can be built from two-level systems by modular construction.

There is associated with each control unit in the system a so-called control problem; the control unit derives the appropriate control by solving its control problem. (In the optimal control approach the problem is one of optimization; namely, the control unit derives a control that gives optimal performance over a class of controls.) The control problem of an infimal unit is referred to as an *infimal problem* and that of the supremal unit is termed a *coordination problem*. The infimal problems in the subsequent sections are assumed to be optimal control problems. However, there is at present no need to so specify the nature of the infimal problems.

In addition to the infimal problems and the coordination problem there is a control problem for the entire system. This control problem is termed an *overall problem*. It is the control problem of the system taken as a whole, integrated unit or the entire control hierarchy viewed as a single control unit. Its solutions ought to specify the set of all controls, i.e., the controls of the entire system ought to represent the solutions of the overall problem.

The overall problem can be used to evaluate the functioning of the control hierarchy and in particular the supremal unit. Assume the overall problem and the family of all infimal problems are given. The supremal unit then has the task of coordinating the infimal units so that the overall control objective as given by the overall problem is achieved.

A relationship among the three types of problems associated with a two-level system is now given formally. The relationship precisely states how coordination can be accomplished under favorable conditions. It also provides a guideline for the synthesis of the supremal unit.

First, the concept of a coordination object for an arbitrary family of control problems is introduced and the law of consistency is stated. Then, the law of consistency is used to express the desired relationship between the three types of control problems associated with a two-level system.

Let $P = \{\Pi_1,\ldots,\Pi_n\}$ be a family of control problems Π_i. An n-tuple $x = (x_1,\ldots,x_n)$ is a *solution of* P if each x_i is a solution of the corresponding control problem Π_i. Let Π denote another control problem. The *solutions of* P *yield solutions of* Π *under a transformation* τ, expressed as

$$(sol\ P) \underset{\tau}{\rightarrow} (sol\ \Pi) \tag{1}$$

if τ takes every solution of P into a solution of Π. If τ is identity then

(1) means that a solution of P is a solution of Π.

Let C be an arbitrary (non-empty) set. The set C is a *coordination object of* P if each $\lambda \in C$ uniquely specifies a family $P_\lambda = \{\Pi_{1\lambda},\ldots,\Pi_{n\lambda}\}$ of control problems and for some $\lambda \in C$, $P_\lambda = P$. Let C be a coordination object for P and for each $i = 1,\ldots,n$ let $P_i \equiv \{\Pi_{i\lambda} : \lambda \in C\}$. Then *the families* $P_1,\ldots,P_n$ *are coordinable relative to* Π *under* τ if there exists a $\lambda^* \in C$ such that

$$(sol\ P_{\lambda^*}) \underset{\tau}{\rightarrow} (sol\ \Pi). \tag{2}$$

To aid in the selection of λ^* a problem Π_0 is formulated. And to provide a guide in formulating Π_0 the relationship

$$[(sol\ P_\lambda) \underset{\tau_0}{\rightarrow} (sol\ \Pi_0)]\ \textit{implies}\ [(sol\ P_\lambda) \underset{\tau}{\rightarrow} (sol\ \Pi)], \tag{3}$$

called the *law of consistency*, is offered.

Consider now how the law of consistency can be used for the synthesis of the supremal (coordination) problem.

Let Π_i be the i-th infimal control problem and let Π be the overall control problem. It should be expected that (1) does not hold in general. The infimal problems are local problems and neglect interactions. Moreover, there is usually a conflict between the infimal and overall objectives. The transformation τ appearing in (1) is usually prespecified. Hence the parameters $\lambda \in C$ are introduced and used to modify the original infimal problems Π_i. The i-th infimal unit then has a family $P_i = \{\Pi_{i\lambda} : \lambda \in C\}$ of control problems rather than the single problem Π_i.

Let Π_0 be the supremal problem. The supremal problem is the coordination problem which is to find $\lambda^* \in C$ and hence appropriate modifications of the infimal control problems so that (2) holds. The law of consistency specifies logical requirements which the supremal, infimal, and overall problems must satisfy for given mappings τ_0 and τ. If for the given τ_0 and τ and each $\lambda \in C$ the law of consistency holds, then the supremal unit need only to concentrate on finding a solution to its problem Π_0. The solution of Π_0 then produces a coordination parameter λ^* that coordinates the infimal control problems relative to the overall control problem Π, i.e., (2) holds.

The law of consistency is trivially obeyed for all $\lambda \in C$ if $\Pi_0 = \Pi$. But then the rational of a multi-level structure is lost, for the supremal unit is required to solve the entire overall problem. The infimal units are then only transmitters or implementers of the control. The art in structuring a two-level system might very well be considered as the selection of the supremal problem Π_0 so that it is considerably simpler (in an appropriate sense) than the overall problem Π and, yet, the law of consistency is obeyed.

Coordination Principles

Suppose a two-level system (as indicated in Figure 1), an overall problem Π, and a family $\{\Pi_1,\ldots,\Pi_n\}$ of infimal problems are given. How should one define the supremal control problem Π_0, the coordination problem, so that the law of consistency (Eq. 3) holds?

The problem confronted here is conceptually very difficult. So far there is no indication how one should proceed. However, in order to attack the problem an approach is postulated that provides a basis for detailed mathematical investigation. The approach consists of two steps: (i) postulating coordination principles that specify the supremal unit's strategy and (ii) investigating the applicability of the coordination principles.

The coordination principles are introduced within the framework of general systems theory, i.e., using set-theoretic notions, in order to present the concepts in a general yet precise form.

Consider a two-level system given as follows:

I. The Overall Problem. Assume there are given two mappings: a process $P : M \to Y$ and a performance $G : M \times Y \to V$ with M the set of controls, Y the set of outputs, and V the set of performance values. The sets M and Y are arbitrary and V is assumed linearly ordered. Define an *overall cost function* $g : M \to V$ such that for all $m \in M$

$$g(m) = G(m,P(m)). \tag{4}$$

The overall problem Π is then to *find* $\hat{m} \in M$ *which minimizes* g on M.

II. The Infimal Problems. The infimal problems are defined in terms of subprocesses obtained from P by decomposition and performance functions defined on the subprocess variables. Let $M = M_1 \times \ldots M_n$ and $Y = Y_1 \times \ldots Y_n$. Decomposition of P, as shown in Figure 2, yields the subprocesses $P_i : M_i \times U_i \to Y_i$ and their interconnections $K_i : M \times Y \to U_i$. The P_i are used to define subprocess models for the infimal problems. The infimal or local performance functions $G_i : M_i \times U_i \times Y_i \to V$ are assumed to be given, and at the outset are assumed to have no specific relationship with G. Define the *infimal cost function* $g_i : M_i \times U_i \to V$ such that for all $m_i \in M_i$ and $u_i \in U_i$

$$g_i(m_i,u_i) = G_i(m_i,u_i,P_i(m_i,u_i)). \tag{5}$$

Two cases arise as to how a means of coordination might be provided.

Case A: Model Coordination. Let $A = U_1 \times \ldots U_n$. Each $\alpha = (\alpha_1,\ldots,\alpha_n) \in A$ is a prediction of the interface inputs and gives for $i = 1,\ldots,n$ the subprocess model $P_{i\alpha}(m_i) = P_i(m_i,\alpha_i)$. Let $C = A$, then for each $\alpha \in A$ the infimal problem $\Pi_{i\alpha}$ is to *find* $\hat{m}_i \in M_i$ *such that*

$$g_i(\hat{m}_i,\alpha_i) = \min_{M_i} g_i(m_i,\alpha_i). \tag{6}$$

Minimization is only over the set M_i of local controls.

Case B: Goal Coordination. Let $\mathcal{B}$ be a given set such that each $\beta \in \mathcal{B}$ specifies for each $i = 1,\ldots,n$ a map $G_{i\beta} : M_i \times U_i \times Y_i \to V$ which is a modification of the original G_i. Let the mappings $g_{i\beta} : M_i \times U_i \to V$ be the infimal cost functions defined in terms of P_i and $G_{i\beta}$ as in (5). Let $C = \mathcal{B}$, then for each $\beta \in \mathcal{B}$ the infimal problem $\Pi_{i\beta}$ is to *find* $(\hat{m}_i,\hat{u}_i) \in M_i \times U_i$ *such that*

$$g_{i\beta}(\hat{m}_i,\hat{u}_i) = \min_{M_i \times U_i} g_{i\beta}(m_i,u_i) \tag{7}$$

Minimization is over both sets M_i and U_i. The interface inputs are treated as free variables.

A coordination principle is postulated for each of the two above cases.

Let the infimal problems be given as in Case A. For each $\alpha \in A$ let $\hat{m}_1(\alpha),\ldots,\hat{m}_n(\alpha)$ be locally optimal, as defined by (6), and let $\hat{m}(\alpha) = (m_1(\alpha),\ldots,m_n(\alpha))$. If $m(\alpha)$ is implemented (applied to the process P) the actual interface input to P_i would be $u_i(\alpha) = K_i(\hat{m}(\alpha),P(\hat{m}(\alpha)))$. Hence the following principle is proposed:

Interaction Prediction Principle

Let $\alpha = (\alpha_1,\ldots,\alpha_n)$ *be the interface inputs predicted and* $u_1(\alpha),\ldots,u_n(\alpha)$ *the actual interface inputs occurring when the locally optimal controls* $m_1(\alpha),\ldots,m_n(\alpha)$ *are applied. The overall optimum is achieved whenever*

$$\alpha_i = u_i(\alpha) \text{ for all } i = 1,\ldots,n \tag{8}$$

i.e., the predicted interface inputs are correct.

Application of this principle is diagrammed in Figure 3. The supremal unit makes a prediction α of the interface inputs, observes the error ε between α and the actual interface inputs $u(\alpha)$, and then corrects or updates its original prediction α to obtain a new prediction α'.

For simplicity the interaction prediction principle is referred to as the INPRE principle. If the INPRE principle applies it yields immediately the supremal problem Π_0 : *find* $\hat{\alpha} = (\hat{\alpha}_1,\ldots,\hat{\alpha}_n) \in A$ *such that* $\varepsilon_i = \hat{\alpha}_i - u_i(\hat{\alpha}) = 0$ *for each* $i = 1,\ldots,n$.

Alternatively if ε_i cannot be made zero one can define an appropriate function f of the errors $\varepsilon_1,\ldots,\varepsilon_n$ and the supremal problem as minimization of that function, i.e., find $\alpha \in A$ such that $f(\varepsilon_1,\ldots,\varepsilon_n)$ attains a minimum.

Consider the Case B. Each of the infimal units is optimizing with respect to both local control and interface inputs. For each $\beta \in B$ and $i = 1,\ldots,n$ let $(\hat{m}_i(\beta),\hat{u}_i(\beta))$ be optimal, as defined in (7). Only the control $\hat{m}(\beta) = (\hat{m}_1(\beta),\ldots,\hat{m}_n(\beta))$ can be applied to the process. Therefore if $\hat{m}(\beta)$ is applied the interface input $u_i(\beta) = K_i(\hat{m}(\beta),P(\hat{m}(\beta))$ appears at P_i.

Interaction Balance Principle

Let $\beta \in B$ be given, let $\hat{u}_1(\beta),\ldots,\hat{u}_n(\beta)$ be the interface inputs as required by the infimal units to achieve local optimum, and let $\hat{u}_1(\beta),\ldots,\hat{u}_n(\beta)$ be the interface inputs that occur if the control $\hat{m}(\beta) = (\hat{m}_1(\beta),\ldots,\hat{m}_n(\beta))$ is actually applied. The overall optimum is achieved whenever

$$\hat{u}_i(\beta) = u_i(\beta) \text{ for each } i = 1,\ldots,n \tag{9}$$

i.e., the actual interface inputs are precisely those required by local optimization.

This principle is referred to as the INBAL principle. If INBAL applies, the problem of the supremal unit is to *find* $\beta \in B$ *such that* $\varepsilon_i = u_i(\beta) - u_i(\beta) = 0$ *for each* $i = 1,\ldots,n$. Or, alternatively, $f(\varepsilon_1,\ldots,\varepsilon_n)$ is minimized, where f is a suitable function of the errors.

Application of INBAL is diagrammed in Figure 4. The supremal unit chooses a coordination parameter $\beta \in B$, observes the error ε between the demanded interface inputs $u(\beta)$ and the interface inputs $u(\beta)$ that actually occur, and then corrects or updates its original selection β to obtain a new coordination parameter β'.

The next question to consider is the application of the principles. Two concepts are used in this context.

INPRE or INBAL is *applicable* if the overall optimum is achieved whenever condition (8) or (9), respectively, is satisfied. There is no requirement here that there actually exist an $\alpha \in A$ satisfying (8) or a $\beta \in B$ satisfying (9), i.e., it is not necessary that the condition making the required error function zero actually exist. However, if the error function vanishes then the overall optimum must result.

The system is *coordinable via* INPRE or INBAL if the principle is applicable and the conditions that yield the vanishing of the required error function exist. That is, in the case of INPRE, the principle must be applicable and there must exist an $\alpha \in A$ satisfying (8); and, in the case of INBAL, the principle must be applicable and there must exist a $\beta \in B$ satisfying (9).

Application of the Coordination Principles

Assume that the sets M, U, and Y are subsets of normed linear spaces while V is the set R of reals, i.e., V = R. This allows the use of methods of functional analysis. Moreover, it provides sufficient mathematical structure so that realistic specifications can be taken into account in sufficient detail. The results, however, remain fairly general. Many of them apply to both dynamic and static process and, furthermore, to non-linear as well as linear processes. In an attempt to highlight the practical implications of the results the following format is followed: a statement of the theorem and then a simple example are given. The theorem statements and following examples contain suitable information for application. Proofs of the theorems are given in the appendix for the sake of verification.

Due to the space limitation only the application of the INBAL principle is considered. Analogous results for the application of the INPRE principle can be found elsewhere[1].

The relationship between the infimal processes $P_1,\ldots,P_n$ and the overall process P is determined since the former are obtained from the latter by decomposition. However, no relationship between the infimal and overall performance functions has been so far assumed, except that they are defined on the same set of variables (controls and outputs). Conditions for applicability and coordinability of the coordination principles are therefore given essentially in terms of the properties of the performance functions on different levels and their interrelationships.

Theorem I. *Let ψ be a mapping $\psi : R^n \to R$ such that for all $\beta \in B$ and all* $m = (m_1,\ldots,m_n) \in M$

$$\psi(g_{1\beta}(m_1,u_1),\ldots,g_{n\beta}(m_n,u_n)) = g(m) \tag{10}$$

where $u_i = K_i(m,P(m))$ for each $i = 1,\ldots,n$. If ψ is order preserving in each of its arguments, then INBAL *is applicable.*

Example I. The simplest example of ψ being order preserving in each of its arguments is

$$\psi(r_1,\ldots,r_n) = a_1r_1+\ldots+a_nr_n;\ a_i \geq 0.$$

If the r_i are restricted to being non-negative, other examples of ψ are polynomials in $(r_1,\ldots,r_n)$ with non-negative coefficients, i.e.,

$$\psi(r_1,r_2,r_3) = a_1r_1^2 + a_2r_1r_2 + a_3r_3;\ a_i \geq 0.$$

The overall cost g is *additive* if for all $m = (m_1,\ldots,m_n) \in M$

$$g(m) = \sum_i g_i(m_i,K_i(m,P(m))). \tag{11}$$

For each $i = 1,\ldots,n$ let f_i be the mapping $f_i : B \times M_i \times U_i \to V$ such that

$$g_{i\beta}(m_i,u_i) = g_i(m_i,u_i) + f_i(\beta,m_i,u_i) \tag{12}$$

on $B \times M_i \times U_i$. Infimal modification is *additive zero-sum* if

$$\sum_i f_i(\beta, m_i, K_i(m, P(m))) = 0 \tag{13}$$

for all $m = (m_1, \ldots, m_n) \in M$ and $\beta \in B$.

Theorem II. *Suppose the overall cost is additive and infimal modification is additive zero-sum modification. Then* INBAL *is applicable.*

Example II. Let n=2 and let M_i, U_i, and Y_i be the set R of reals. Let the interconnections be $u_1 = y_2$ and $u_2 = y_1$, i.e., the maps K_i are projection mappings. Define G such that

$$G(m_1, m_2, y_1, y_2) = G_1(m_1, u_1, y_1) + G_2(m_2, u_2, y_2)$$

where $u_1 = y_2$ and $u_2 = y_1$. This gives additive overall cost:

$$g(m_1, m_2) = g_1(m_1, u_1) + g_2(m_2, u_2)$$

whenever $(u_2, u_1) = (y_1, y_2) = P(m_1, m_2)$. Let $B = R \times R$ and for each $\beta = (\beta_1, \beta_2) \in B$ define

$$G_1\ (m_1, u_1, y_1) \equiv G_1(m_1, u_1, y_1) + \beta_1 u_1^2 - \beta_2 y_1^2$$

and

$$G_2\ (m_2, u_2, y_2) \equiv G_2(m_2, u_2, y_2) + \beta_2 u_2^2 - \beta_1 y_2^2 .$$

This is additive zero-sum infimal modification, i.e., the f_i in (12) are:

$$f_1(\beta, m_1, u_1) = \beta_1 u_1^2 - \beta_2 (P_1(m_1, u_1))^2$$

and

$$f_2(\beta, m_2, u_2) = \beta_2 u_2^2 - \beta_1 (P_2(m_2, u_2))^2.$$

Theorem II states that the INBAL principle is applicable to this system.

In the following theorems we make use of the mapping $\nu : B \times M \times U \to V$, where $U = U_1 \times \ldots U_n$, defined by the equation:

$$\nu(\beta, m, u) = \sum_i g_{i\beta}(m_i, u_i) \tag{14}$$

for each $\beta \in B$, $m = (m_1, \ldots, m_n) \in M$, and $u = (u_1, \ldots, u_n) \in U$.

Theorem III. *Suppose the following hold:*

(i) *the overall cost* g *is additive and has a minimum on* M;

(ii) *infimal modification is additive zero-sum modification;*

(iii) *there is a non-empty subset* $B^0 \subseteq B$ *such that for all* $\beta \in B^0$ *each infimal cost* $g_{i\beta}$ *admits a unique minimum on* $M_i \times U_i$;

(iv) *for some* $\hat{\beta} \in B^0$, $\inf_{M \times U} \nu(\beta, m, n) = \inf_M g(m)$.

Then the two-level system is coordinable via INBAL.

Example III. In Example II define

$$y_1 = P_1(m_1,u_1) \equiv 2m_1 + u_1$$

and

$$y_2 = P_2(m_2,u_2) \equiv 2m_2 - u_2.$$

Define $G_i(m_i,u_i,y_i) \equiv m_i^2 + (y_i-1)^2$ for $i = 1,2$. Hence for $\beta = (\beta_1,\beta_2) \in R \times R$, the modified infimal costs are

$$g_{1\beta}(m_1,u_1) = m_1^2 + (2m_1 + u_1-1)^2 + \beta_1 u_1^2 - \beta_2(2m_1 + u_1)^2$$

and

$$g_{2\beta}(m_2,u_2) = m_2^2 + (2m_2 - u_2-1)^2 + \beta_2 u_2^2 - \beta_1(2m_2 - u_2)^2.$$

Define $\mathcal{B}^0 \subseteq \mathcal{B}$ such that $\mathcal{B}^0$ contains all $\beta = (\beta_1,\beta_2) \in \mathcal{B}$ for which $\beta_1 \neq (5\beta_2 + 1)(4\beta_2 + 1)^{-1}$ and $\beta_2 \neq (5\beta_1 + 1)(4\beta_1 + 1)^{-1}$. Then, for all $\beta \in \mathcal{B}^0$ each infimal cost $g_{i\beta}$, $i = 1,2$, admits a unique minimum on $M_i \times U_i$. Furthermore, for $\hat{\beta} = (0, -1/2) \in \mathcal{B}^0$,

$$\min_{M\times U} [g_{1\hat{\beta}}(m_1,u_1) + g_{2\hat{\beta}}(m_2,u_2)] = \frac{2}{3} = \min_{M} g(m_1,m_2).$$

The conditions of Theorem III are satisfied and, therefore, the system described is coordinable via INBAL. In fact, for $\beta = (0, -1/2)$ we have $(\hat{m}_1(\hat{\beta}),\hat{u}_1(\hat{\beta})) = (0,2/3)$ and $(\hat{m}_2(\hat{\beta}),\hat{u}_2(\hat{\beta})) = (2/3,2/3)$ and, moreover, $\hat{m}(\beta) = (0, 2/3)$ minimizes $g(m_1,m_2)$ on $M_1 \times M_2$.

Let the mapping $K : M \times U \to U$ be such that for all $m = (m_1,\ldots,m_n) \in M$ and $u = (u_1,\ldots,u_n) \in U$

$$K(m,u) = (K_1(m,y),\ldots,K_n(m,y)) \tag{15}$$

where $y = (P_1(m_1,u_1),\ldots,P_n(m_n,u_n))$. Infimal modification is *linear difference modification* if $\mathcal{B}$ is the conjugate space of U and for all $\beta \in \mathcal{B}$, $m = (m_1,\ldots,m_n) \in M$, and $u = (u_1,\ldots,u_n) \in U$

$$\sum_i f_i(\beta,m_i,u_i) = \beta(u - K(m,u)). \tag{16}$$

Theorem IV. *Suppose the following hold:*

(i) *the overall cost* g *is additive and has a minimum on* M*;*

(ii) *infimal modification is linear difference modification;*

(iii) *each* $g_{i\beta}$ *is strictly convex and lower semi-continuous on* $M_i \times U_i$ *and each set* $M_i \times U_i$ *is weakly compact and convex;*

(iv) *for some* $\beta \in \mathcal{B}$, $\sup_{\mathcal{B}} \inf_{M\times U} \nu(\beta,m,u) = \inf_{M\times U} \nu(\beta,m,u)$.

Then the two-level system is coordinable via INBAL.

Example IV. Let n=2, let Y_i be the set R of reals, and let M_i and U_i be subsets of R:

$$M_i \equiv \{m_i \in R : |m_i| \leq 1\},\ i = 1,2;$$

and

$$U_i \equiv \{u_i \in R : |u_i| \leq r\},\ i = 1,2$$

where r is given such that $1 < r < \frac{10}{3}$. Define

$$y_1 = P_1(m_1,u_1) \equiv (m_1 - 1)^2 + u_1^2$$

and

$$y_2 = P_2(m_2,u_2) \equiv u_2(m_2 + 2)$$

and let the equation $u_1 = m_2$ and $u_2 = m_1$ define the interconnections, i.e., the K_i are projection mappings. Define

$$G_1(m_1,u_1,y_1) \equiv y_1 + m_1^2$$

$$G_2(m_2,u_2,y_2) \equiv y_2^2 + 10m_2^2$$

and

$$G(m_1,m_2,y_1,y_2) \equiv G_1(m_1,u_1,y_1) + g_2(m_2,u_2,y_2)$$

The infimal costs are then

$$g_1(m_1,u_1) = (m_1 - 1)^2 + u_1^2 + m_1^2$$

and

$$g_2(m_2,u_2) = u_2^2(m_2 + 2)^2 + 10m_2^2.$$

Moreover they are strictly convex and bounded on the sets $M_i \times U_i$ respectively. The overall cost

$$g(m_1,m_2) = m_1^2 + (m_1 - 1)^2 + m_1^2(m_2 + 2)^2 + 11m_2^2$$

is additive and strictly convex on $M_1 \times M_2$ and therefore has a minimum on $M_1 \times M_2$. Let the f_i in (12) be such that

$$f_1(\beta,m_1,u_1) = \beta_1 u_1 - \beta_2 m_1$$

and

$$f_2(\beta,m_2,u_2) = \beta_2 u_2 - \beta_1 m_2$$

where $\beta = (\beta_1,\beta_2) \in R \times R$. Hence, for $B = R \times R$, infimal modification is linear difference modification and for all $\beta \in B$ each modified infimal cost $g_{i\beta}$ is strictly convex on $M_i \times U_i$. With some effort one can find a $\beta \in B$ such that condition (iv) of Theorem IV is satisfied. Therefore according to Theorem IV the system described in coordinable via INBAL. In fact $\hat{\beta}$ satisfies condition (9).

Theorem V. *Suppose the following hold:*

(i) *the overall cost* g *is additive and has a minimum on* M*;*

(ii) *infimal modification is linear difference modification;*

(iii) *each* $g_{i\beta}$ *is bounded, strictly convex, and lower semi-continuous on*

$M_i \times U_i$ *and the sets* M_i *and* U_i *are closed, bounded and convex;*

(iv) $0 \in \mathrm{Int}(E)$ *where* $E \equiv \{e = u - K(m,u): m \in M$ *and* $u \in U\}$.

Then the two-level system is coordinable via INBAL.

Example V. Consider the standard linear dynamic system with quadratic cost problem. For $i = 1,\ldots,n$ let E_{im} and E_{iy} be Euclidean spaces and suppose the subprocesses P_i are defined for $t \in [0,1]$ by the linear vector-matrix equations

$$\dot{y}_i(t) = A_i y_i(t) + B_i m_i(t) + u_i(t);\ i=1,\ldots,n \tag{17}$$

where $y_i(t)$ and $u_i(t) \in E_{iy}$ and $m_i(t) \in E_{im}$. The matrices A_i and B_i are of appropriate dimension and may be time varying but continuous on $[0,1]$. Let M_i be the set of $m_i \in \mathcal{L}_2(0,1)$ such that $m_i(t) \in \Omega_i$ a compact convex subset of E_{im}. Define E_{iy} to be the set of $e_i \in \mathcal{L}_2(0,1)$ such that $e_i(t) \in E_{iy}$ and $||e_i(t)|| \leq \varepsilon$ for a given $\varepsilon > 0$.

Suppose the equations

$$u_i(t) = \sum_j A_{ij} y_j(t)\ ;\ A_{ii} \equiv 0; i=1,\ldots,n \tag{18}$$

define the interconnections. Let $y_i(0) \in E_{iy}$ be given for each $i = 1,\ldots,n$. Define Y as the set of solutions $y = (y_1,\ldots,y_n)$ of

$$\dot{y}_i(t) = A_i y_i(t) + B_i m_i(t) + \sum_j A_{ij} y_j(t) + e_i(t);\ i=1,\ldots,n \tag{19}$$

for $m_i \in M_i$ and $e_i \in E_{iy}$. Y is bounded and convex. Therefore, for $i=1,\ldots,n$ let U_i be a closed, bounded, and convex set such that

$$\sum_j A_{ij} Y_j \subseteq U_i$$

where Y_j is the j-th coordinate of Y. The sets U_i then contain all possible interface inputs.

Define the infimal performance functions

$$G_i(m_i,u_i,y_i) \equiv \int_0^1 c_i(m_i(t),y_i(t))dt;\ i = 1,\ldots,n$$

where the c_i are positive definite quadratic forms. Let the overall performance function G be given as the sum of the G_i. When the subprocesses, given by (17), are interconnected via (18) the overall cost has a minimum.

Let $\mathcal{B} = U_1^* \times \ldots \times U_n^*$ where each U_i^* is the conjugate space of U_i. For each $\beta = (\beta_1,\ldots,\beta_n)$ let the modified infimal performance function be

$$G_{i\beta}(m_i,u_i,y_i) \equiv G_i(m_i,u_i,y_i) + \beta_i(u_i) - \sum_j \beta_j(A_{ji} y_i).$$

Infimal modification is then linear difference modification.

All the conditions of Theorem V are satisfied by the system described. In particular, since M_i and U_i are bounded it follows from (19) and the continuity of c_i that condition (iv) holds. Therefore the system is coordinable via INBAL.

Concluding Remarks

The prime reason for the formulation of the coordination principles is the solution of the structural problem in the synthesis of multi-level systems. The framework of optimal control is used only as a convenience to show mathematically that the principles indeed do apply to a broad class of situations. However, the importance of the principles lies beyond the limited framework of optimal control. Indeed, the coordination principles define the supremal problem as a prediction and regulation problem. Using the INPRE principle, the supremal has to predict the interface inputs, compare them with the actual interface inputs, and update the prediction whenever the error is evaluated as being outside acceptable bounds. Similar interpretation can be given to the INBAL principle. Using the INBAL principle the supremal has to modify the infimal performance functions, compare the interface inputs demanded by the infimal units and those that actually occur, and provide a new modification of the infimal performance functions whenever the error is found to be outside acceptable bounds. In such a way the principles can be used to synthesize practical supervisory control algorithms that can be applied on-line to control the systems in the presence of disturbances. For example, if the INPRE principle is applicable, all that is needed to decentralize the control action is an acceptably accurate prediction of interface inputs. This method can therefore be used in the large scale multi-variable control situations where the problem of dimensionality is overwhelming.

Acknowledgment

The research leading to this paper was supported in part by ONR 1141(13), NSF GK1394 and NSF GK3396. Many of our co-workers of the Systems Research Center have contributed to the research reported here. Detailed account of these and other historical references are given in [1].

Reference

[1] M.D. Mesarovic, D. Macko, Y. Takahara, Theory of Multi-Level Systems, (book to be published).

Appendix A

Proof of Theorem I

Suppose ψ is order preserving in each of its arguments. Let $\beta \varepsilon \mathcal{B}$ be given and for each $i = 1,\ldots,n$ suppose $(\hat{m}_i(\beta),\hat{u}_i(\beta))$ is optimal as defined by (7) and $\hat{u}_i(\beta) = K_i(\hat{m}(\beta),P(\hat{m}(\beta)))$, where $\hat{m}(\beta) = (\hat{m}_1(\beta),\ldots,\hat{m}_n(\beta))$. To prove applicability of INBAL it is shown that $\hat{m}(\beta)$ minimizes g over M.

Let $m = (m_1,\ldots,m_n) \in M$ be arbitrary and for each $i = 1,\ldots,n$ let $u_i = K_i(m,P(m))$. Then $g_{i\beta}(\hat{m}_i(\beta),\hat{u}_i(\beta)) \le g_{i\beta}(m_i,u_i)$ for each $i = 1,\ldots,n$. Therefore, since ψ satisfies (10) and is order preserving in each of its arguments,

$$g(\hat{m}(\beta)) = \psi(g_{1\beta}(\hat{m}_1(\beta),\hat{u}_1(\beta)),\ldots,g_{n\beta}(\hat{m}_n(\beta),\hat{u}_n(\beta)))$$
$$\le \psi(g_{1\beta}(m_1,u_1),\ldots,g_{n\beta}(m_n,u_n)) = g(m).$$

Hence $\hat{m}(\beta)$ minimizes g over M. This completes the proof.

<u>Proof of Theorem II</u>.

If g is additive and the infimal modification is additive zero-sum modification then (11), (12), and (13) are always satisfied. Hence for all $\beta \in B$ and $m = (m_1,\ldots,m_n) \in M$

$$\sum_i g_{i\beta}(m_i,K_i(m,P(m))) = \sum_i g_i(m_i,K_i(m,P(m))) = g(m).$$

Therefore, by Theorem 1, INBAL is applicable. This completes the proof.

<u>Proof of Theorem III</u>.

Because of (i) and (ii) it follows from Theorem II that INBAL is applicable. It remains to be shown that for some $\beta \in B$ condition (9) is satisfied. It is then shown that $\beta = \hat{\beta}$ in (iv) is such a β. Because of (i) and (iii), condition (iv) hold for minima as well as infima. Let $(\hat{m}_i(\beta),\hat{u}_i(\beta))$, for $i = 1,\ldots,n$, be optimal, as defined by (7), and let $\hat{m} = (\hat{m}_1,\ldots,\hat{m}_n)$ minimize g over M. Then from (iv) and since g is additive,

$$\sum_i g_{i\beta}(\hat{m}_i(\beta),\hat{u}_i(\beta)) = g(\hat{m}) = \sum_i g_{i\beta}(\hat{m}_i,K_i(\hat{m},P(\hat{m}))).$$

Hence, for each $i = 1,\ldots,n$

$$g_{i\beta}(\hat{m}_i(\beta),\hat{u}_i(\beta)) = g_{i\beta}(\hat{m}_i,K_i(\hat{m},P(\hat{m})).$$

Therefore, from (iii), $\hat{m}_i(\beta) = \hat{m}_i$ and $\hat{u}_i(\beta) = K_i(\hat{m},P(\hat{m}))$ for each $i = 1,\ldots,n$. This completes the proof.

<u>Proof of Theorem IV</u>.

It suffices to show that conditions III-(ii), III-(iii) and III-(iv) hold.

It follows from (15) and (16) that linear difference infimal modification is also additive zero-sum infimal modification. Hence condition III-(ii).

Each $g_{i\beta}$ is strictly convex and lower semi-continuous on $M_i \times U_i$ and $M_i \times U_i$ is weakly compact and convex. Hence $g_{i\beta}$ has a unique minimum on $M_i \times U_i$. Hence condition I I-(iii).

Condition III-(iv) is now shown to hold. For any $m \in M$ and $\beta \in B$, because of (i) and III-(ii), i.e., equations (12) and (13), $\nu(\beta,m,K(m,P(m))) = g(m)$ and hence

$$\sup_{B} \nu(\beta,m,K(m,P(m))) = g(m).$$

Let $m = (m_1,\ldots,m_n) \in M$ and $u = (u_1,\ldots,u_n) \in U$. Then from (12), (13), (14), and (16),

$$\sup_{B} \nu(\beta,m,n) = \sum_{i} g_i(m_i,u_i) + \sup_{B} \beta(u - K(m,u)).$$

From the Hahn-Banach theorem, $\sup_{B} \beta(u - K(m,u)) = \infty$ if $u \neq K(m,u)$. Therefore, for all $m \in M$ and $u \in U$,

$$\sup_{B} \nu(\beta,m,u) = \begin{cases} g(m) \text{ if } u = K(m,P(m)), \\ \infty \quad \text{otherwise} \end{cases}$$

Hence

$$\inf_{M\times U} \sup_{B} \nu(\beta,m,u) = \inf_{M} g(m) \qquad \text{(A-1)}$$

As a result of condition (ii) and (iii), ν is weakly lower semi-continuous and convex on $M \times U$ for each $\beta \in B$ and, moreover, ν is concave on B for each $m \in M$ and $u \in U$. Therefore,

$$\sup_{B} \inf_{M\times U} \nu(\beta,m,u) = \inf_{M\times U} \sup_{B} \nu(\beta,m,u). \qquad \text{(A-2)}$$

Due to (iv) there exists $\hat{\beta} \in B$ such that

$$\sup_{B} \inf_{M\times U} \nu(\beta,m,u) = \inf_{M\times U} \nu(\hat{\beta},m,n). \qquad \text{(A-3)}$$

Therefore, from (A-1), (A-2), and (A-3), $\inf_{M\times U} \nu(\hat{\beta},m,u) = \inf_{M} g(m)$ which is condition III-(iv). This completes the proof.

Proof of Theorem V.

It suffices to show that IV-(iii) and IV-(iv) hold.

Note that in a reflexive Banach space a closed bounded and convex set is weakly compact. Therefore condition IV-(iii) holds.

For each $\beta \in B$ since the $g_{i\beta}$ have minima let $\nu^*(\beta) = \inf_{M\times U} \nu(\beta,m,u)$. If ν^* has a maximum on B then condition IV-(iv) is satisfied. According to Appendix B it is sufficient to show that ν^* is continuous and concave on B and $\nu^*(\beta) \to -\infty$ as $||\beta|| \to \infty$.

To show concavity let $\beta,\beta' \in B$. Then because of linearity in β, condition (ii),

$$\nu(\lambda\beta + (1 - \lambda)\beta',m,u) = \lambda\nu(\beta,m,u) + (1 - \lambda)\nu(\beta',m,u)$$

for all $m \in M$, $u \in U$, and any real number λ. Therefore, for

$0 \leq \lambda \leq 1$, one has

$$\nu^*(\lambda\beta + (1 - \lambda)\beta') \geq \lambda\nu^*(\beta) + (1 - \lambda)\nu^*(\beta'),$$

i.e., ν^* is concave.

Continuity follows from the face that a convex functional f on a linear topological space X is continuous on $Y \subseteq X$ if and only if there is a non-empty open set $O \subseteq Y$ such that f(x) is majored by some number α for any $x \in O$. Let $\beta \in B$, $m \in M$, and $u \in U$. Since the U is bounded, K is also bounded since its range is included in U and $|\beta(u - K(m,u))| \leq a\,||\beta||$ for $0 < a < \infty$. Therefore from (12), (13), (14), and (16)

$$\nu(\beta,m,u) \geq \sum_i g_i(m_i,u_i) - a\,||\beta||\,,$$

where $m = (m_1,\ldots,m_n)$ and $u = (u_1,\ldots,u_n)$; and hence

$$\nu^*(\beta) \geq \nu^*(0) - a\,||\beta||\,,$$

i.e., $-\nu^*(\beta)$ is majored in any bounded open set of B. Therefore ν^* is continuous on B.

To show $\nu^*(\beta) \to -\infty$ as $||\beta|| \to \infty$ it is proven that there is an $r > 0$ such that $\nu^*(\beta) < \nu^*(0)$ for every $\beta \in B$ with norm r.

First assume there is an $r > 0$ such that $\nu^*(\beta) < \nu^*(0)$ for every $\beta \in B$ with norm r, and show that this implies that $\nu^*(\beta) \to -\infty$ as $||\beta|| \to \infty$. Let $\beta, \beta' \in B$ and $\beta'' = \lambda\beta + (1 - \lambda)\beta'$. For $0 < \lambda < 1$ one has from concavity

$$\nu^*(\beta'') \geq \lambda\nu^*(\beta' + \tfrac{1}{\lambda}(\beta'' - \beta')) + (1 - \lambda)\nu^*(\beta')$$

by using $\beta = \beta' + \frac{1}{\lambda}(\beta'' - \beta')$. For $\mu = 1/\lambda \geq 1$,

$$\mu\,[\nu^*(\beta'') - \nu^*(\beta')] + \nu^*(\beta') \geq \nu^*(\beta' + \mu(\beta'' - \beta')).$$

Let $||\beta''|| = r$ and $\beta = \mu\beta''$ and $\beta' = 0$. Then for $|\mu| = ||\beta||/r \geq 1$,

$$\frac{||\beta||}{r}\,[\nu^*(\beta'') - \nu^*(0)] + \nu^*(0) \geq \nu^*(\beta).$$

Since $||\beta''|| = r$, $\nu^*(\beta'') < \nu^*(0)$. Therefore $\nu^*(\beta) \to -\infty$ as $||\beta|| \to \infty$.

Now exhibit an $r > 0$ such that $\nu^*(\beta) < \nu^*(0)$ for every $\beta \in B$ with norm r. From (iii) each g_i is bounded on $M_i \times U_i$, hence there exist finite numbers b_1 and b_2 such that for all $m \in M$ and $u \in U$, $b_1 \leq \nu(0,m,u) \leq b_2$. Observe that $\nu(0,m,u) = \sum_i g_i(m_i,u_i)$ where $m = (m_1,\ldots,m_n)$ and $u = (u_1,\ldots,u_n)$. Let $\beta \in B$ be arbitrary. Since for each $m \in M$ and $u \in U$,

$$\nu(\beta,m,u) = \nu(0,m,u) + \beta(u - K(m,u)),$$

one has

$$\nu^*(\beta) \leq b_2 + \inf_{M \times U} \beta(u - K(m,u)).$$

From condition (iv) it follows that there exists an $\varepsilon > 0$ such that

$$\inf_{M\times U} \beta(u - K(m,u)) = \inf_{E} \beta e \leq \inf_{||e||=\varepsilon} \beta e$$

where $E = \{e = u - K(m,u) \mid m \in M \text{ and } u \in U\}$. On the other hand $||\beta|| = \sup_{||e||=1} |\beta e| = \sup_{||e||=1} \beta e = -\inf_{||e||=1} \beta e$. Therefore

$\inf_{||e||=1} \beta e = -||\beta||$. Let $r = b_2 - b_1 + 1$. Then for any $\beta \in B$ with norm $\frac{r}{\varepsilon}$,

$$\begin{aligned} \nu^*(\beta) &\leq b_2 + \inf_{M\times U} \beta(u - K(m,u)) \leq b_2 - \varepsilon\, ||\beta|| \\ &= b_1 - 1 < b_1 \leq \inf_{M\times U} \nu(0,m,u) = \nu^*(0). \end{aligned}$$

Hence, for $\varepsilon||\beta|| = b_2 - b_1 + 1$, $\nu^*(\beta) < \nu^*(0)$. This completes the proof.

Appendix B

Let X be a reflexive Banach space and suppose f is a lower semi-continuous convex functional on X such that $f(x) \to \infty$ as $||x|| \to \infty$. Then f has a minimum on X.

Proof. Since the functional f is lower semi-continuous and convex it is weakly lower semi-continuous. Let $\inf f(x) = \alpha < \infty$, then there exists a sequence $\{x_n\}$ in X such that $f(x_n) \to \alpha$ as $n \to \infty$. Then there exists $r > 0$ and N such that for $n > N$, $||x_n|| < r$. If not, there is a subsequence $\{x_{u'}\}$ of $\{x_n\}$ such that $||x_{n'}|| \to \infty$. Then $f(x_{n'}) \to \infty$ as $n' \to \infty$. But this is a contradiction of $f(x_n) \to \alpha < \infty$ as $n \to \infty$. Let $Y = \{x \in X \mid \; ||x|| \leq r\}$. Then Y is closed, bounded, and convex, that is, weakly compact. Hence there is a subsequence $\{X_{n''}\}$ of $\{X_n\}$ that converges weakly to $x_0 \in Y$. Hence $\alpha \geq \liminf f(x_{n''}) \geq f(x_0) \geq \alpha$ by weak semi-continuity. This completes the proof.

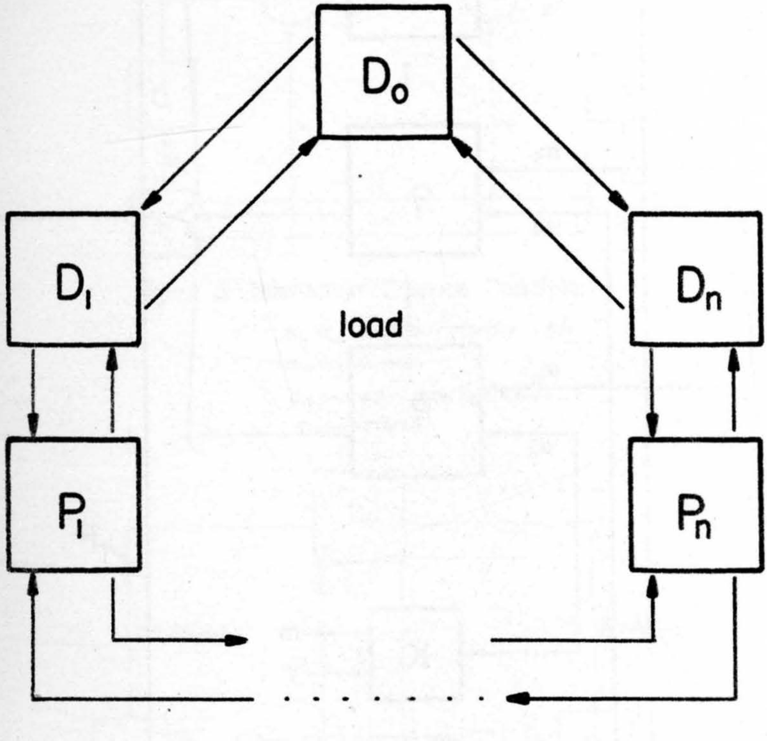

Figure 1. Two-level System

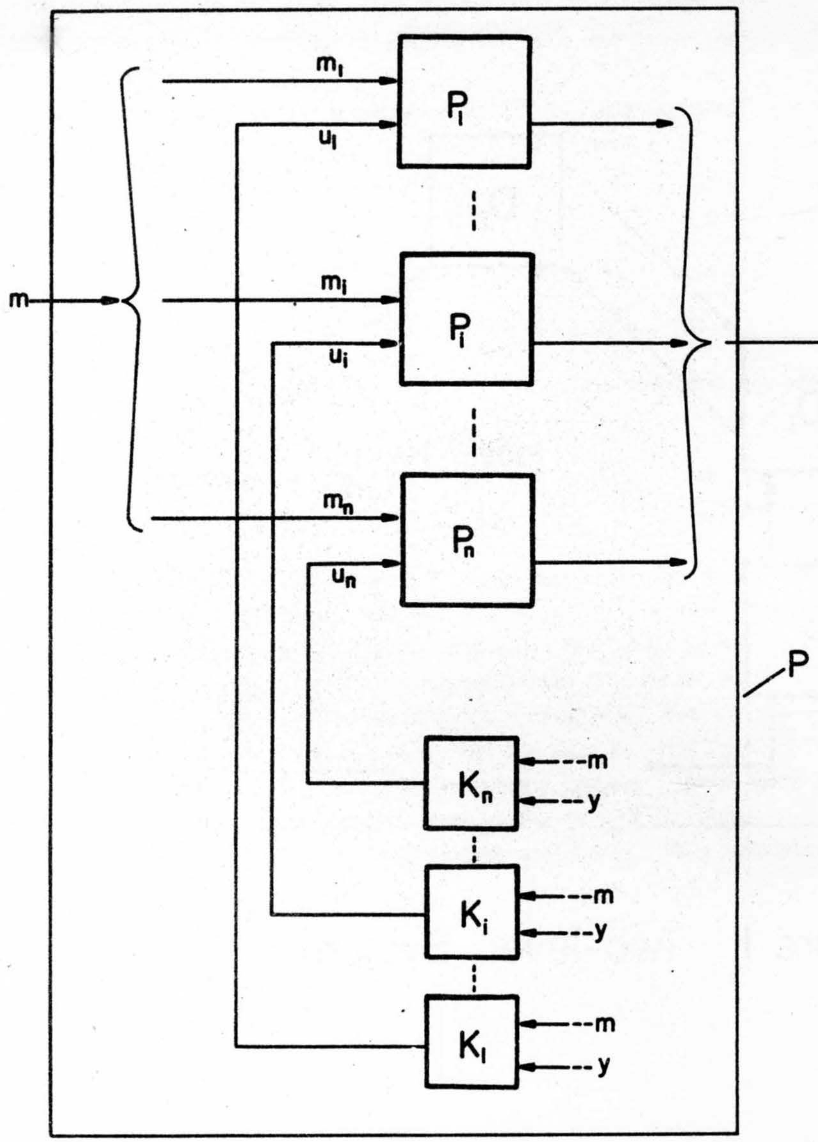

Figure 2. Decomposition of P into Subprocesses $P_1,\ldots,P_n$.

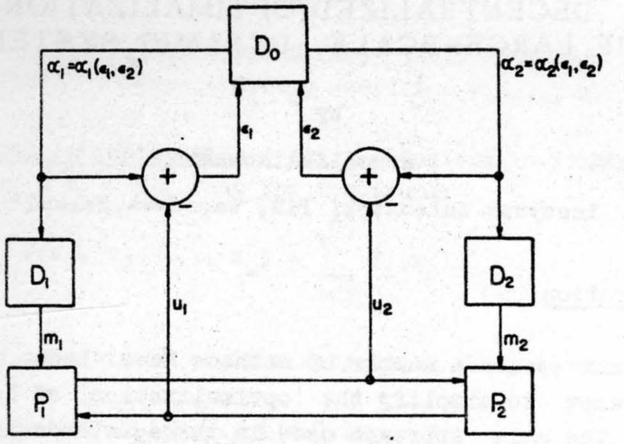

Figure 3. Interaction Balance Principle

α_1, α_2 - predicted interface inputs

m_1, m_2 - controls

u_1, u_2 - actual interface inputs

ϵ_1, ϵ_2 - errors

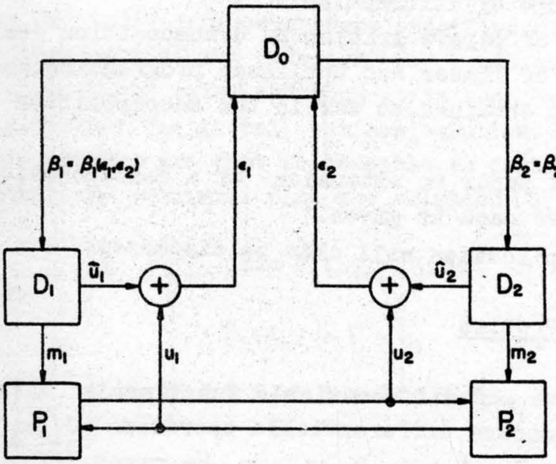

Figure 4. Interaction Balance Principle

β_1, β_2 coordination

$\hat{u}_1, \hat{u}_2$ required interface inputs

m_1, m_2 control

u_1, u_2 actual interface inputs

ϵ_1, ϵ_2 errors

DECENTRALIZED OPTIMALIZATION OF LARGE-SCALE, DYNAMIC SYSTEMS

by

Roman Kulikowski

Instytut Automatyki PAN, Warszawa,Poland

1. Introduction

In recent years a number of methods have been developed which attempt to simplify the optimalization of large scale systems. The main approach used in these methods is based on the concept of decentralization or decomposition, which attempts to break down the large-scale problem into a number of simpler sub-problems. As a result the complex system can be controlled by several controllers, which are arranged in the form of a two-level structure and which are coordinated by a process of exchange of information [1-5, 10].

The majority of papers written on decomposition deals with complex problems of linear and nonlinear programming.Much less results have been obtained,so far,in the decomposition of dynamic problems.

In the present paper an extension of a decomposition method for the dynamic case is given.

A concrete application will also be discussed.

2. Statement of Problem

Let 2m concave and diffrenetiable functionals $F_i(x_i)$ and $H_i(x_i)$ and m concave,differentiable operators $G_i(x_i)$, where $x_i \in X_i$, $G_i(x_i) \in Z_i$, $i = 1, 2, \ldots, m$, be given. X_i, Z_i are assumed to be, in general, real Banach spaces.

The local optimalization problem consists in finding such elements $x_i = \bar{x}_i \in X_i$, that the functionals $F_i(x_i)$ attain their conditional maximum subject to the inequality constraints $G_i(x_i) \geqslant 0$, i.e.

$$F_i(\bar{x}_i) = \max F_i(x_i)$$
$$G_i(x_i) > 0, \quad i = 1, \ldots, m. \qquad (1)$$

The global optimalization problem consists in finding such elements $x_i = \bar{\bar{x}}_i \in X_i$, that the functional

$$F(x_1, x_2, \ldots, x_m) = \sum_{i=1}^{m} F_i(x_i) \qquad (2)$$

attains its conditional maximum subject to the inequality constraints

$$G_i(x_i) \geqslant 0, \quad i = 1, \ldots, m\,, \qquad (3)$$

$$\sum_{i=1}^{m} H_i(x_i) \geqslant h\,, \qquad (4)$$

where h is a given real number.

The relation (4) represents the (functional) interactions among the m individual subsystems specified by F_i and G_i.

It is assumed that the solutions for local and global optimalization problems exist and that it is such easier to derive the local, rather the global, optimum solution. We assume also that there exists, and that is possible,to derive the solution of m auxiliary problems: find the solution $x_i = \tilde{x}_i$ of

$$\max F_i(x_i) \qquad (5)$$

subject to

$$G_i(x_i) \geqslant 0\,, \qquad (6)$$

$$H_i(x_i) = y_i, \quad i = 1, \ldots, m, \qquad (7)$$

where y_i - given real numbers.

These problems shall be called the I-level optimalization problems.

In the case where it is possible to derive $\tilde{x}_i$ as a function of y_i, i. e. $\tilde{x}_i(y_i)$ it is also possible to derive the functions

$$f_i(y_i) = F_i\big[\tilde{x}_i(y_i)\big]\,, \quad i = 1, \ldots, m\,.$$

By II-level optimalization problem we shall call the problem of finding values $y_i = \tilde{y}_i$, $i = 1, \ldots, m$, maximizing the function

$$\sum_{i=1}^{m} f_i(y_i) \tag{8}$$

subject to the constraint

$$\sum_{i=1}^{m} y_i \geqslant h \; . \tag{9}$$

That is a typical nonlinear (statical) optimalization problem.

There exist many known computational techniques such as the gradient projection method [9] which enable the numerical computation of $\tilde{y}_i$ and then $\tilde{x}_i = \tilde{x}_i(\tilde{y}_i)$, $i = 1, \ldots, m$.

In many practical problems,however, it is difficult tosolve the dynamic I-level problems and to derive the explicit form of $f_i(y_i)$.

The main idea of the present paper is to show that the solution $\tilde{y}_i$ can be computed without the knowledge of $f_i(y_i)$. We shall show that it is possible to derive the numerical values of gradient $f_i'(y_i)$, $i = 1, \ldots, m$, for the specified values of y_i and use the gradient technique. In the case of nonlinear differentiable functions such a method has already been used succesfully [10]. The present paper shows that a similar approach can be used also in the case of nonlinear differentiable functionals.

3. Effective Solution of II-level Optimalization Problems

In order to solve the problem (5), (6), (7) we should investigate the generalized lagrangean:

$$\phi_i(x_i, \lambda_i, \mu_i) = F_i(x_i) + \lambda_i \left[G_i(x_i)\right] + \mu_i \left[H_i(x_i) - y_i\right] , \tag{10}$$

where λ_i - linear functionals in the space Z^* (adjoint to Z), μ_i - Lagrange multipliers.

Denote $H_i(\tilde{x}_i)$ by $\tilde{y}_i$ and $F_i(\tilde{x}_i)$ by $\tilde{f}_i$. Assuming that these functionals are continuous, strongly differentiable [x)] and $\|\operatorname{grad} H_i(x_i)\| \neq 0$, $i = 1, \ldots, m$, we shall prove that

$$\frac{d\tilde{f}_i}{d\tilde{y}_i} = \mu_i \ , \quad i = 1, \ldots, m \ . \tag{11}$$

Indeed, we have

$$\tilde{f}_i = \lim_{\gamma \to 0} F_i(\tilde{x}_i + \gamma h_i), \ \tilde{y}_i = \lim_{\gamma \to 0} H_i(\tilde{x}_i + \gamma h_i)$$

where h_i are arbitrary elements of X_i, and γ is a number. Then

$$\frac{d\tilde{f}_i}{d\gamma} = \frac{d}{d\gamma}\left[\lim_{\gamma \to 0} F_i(\tilde{x}_i + \gamma h_i)\right] =$$

$$= \lim_{\Delta\gamma \to 0} \frac{1}{\Delta\gamma}\left\{\lim_{\gamma \to 0} F_i\left[\tilde{x}_i + (\gamma + \Delta\gamma)h_i\right] + \right.$$

$$\left. - \lim_{\gamma \to 0} F_i\left[\tilde{x}_i + \gamma h_i\right]\right\} = \lim_{\Delta\gamma \to 0} \frac{1}{\Delta\gamma}\left\{F_i\left[\tilde{x}_i + \Delta\gamma \cdot h_i\right] + \right.$$

$$\left. - F_i\left[\tilde{x}_i\right]\right\} = dF_i\left[\tilde{x}_i, h_i\right] ,$$

where $dF_i\left[\tilde{x}_i, h_i\right]$ is the strong differential of $F_i\left[\tilde{x}, h\right]$ at the point $\tilde{x}_i \in X_i$ with respect to the variation $h_i \in X_i$.

We have also

---

x) The functional $F(x)$ is called strongly differentiable at x when the limit

$$\lim_{\gamma \to 0} \frac{1}{\gamma}\left[F(x + \gamma h) - F(x)\right] = dF(x, h), \ h \in X$$

exists and $dF(x, h)$ is continuous at x [11].

The strong differential $dF(x, h)$ can be written in the form of a linear functional, i. e. $(f(x), h)$, where $f(x)$ is called strong gradient of $F(x)$, i.e. $f(x) = \operatorname{grad} F(x)$ [11].

$$\frac{d\tilde{f}_i}{d\tilde{y}_i} = \frac{d\tilde{f}_i}{d\gamma}\frac{d\gamma}{d\tilde{y}_i} = dF_i\left[\tilde{x}_i,\ h_i\right]\frac{d\gamma}{dy_i} \tag{12}$$

By differentiation of $\tilde{y}_i = H_i(x_i)$ we find in a similar way: $\frac{dy_i}{d\gamma} = dH_i(\tilde{x}_i,\ h_i)$. That can be written:

$$1 - dH_i(\tilde{x}_i,\ h_i)\,\frac{d\gamma}{d\tilde{y}_i} = 0\ . \tag{13}$$

Multiplying (13) by μ_i and summing up with (12) we get

$$\frac{d\tilde{f}_i}{d\tilde{y}_i} = \left[dF_i(\tilde{x}_i,\ h_i) - \mu_i\ dH_i(\tilde{x}_i,\ h_i)\right]\frac{d\gamma}{d\tilde{y}_i} + \mu_i\ ,$$

$$i = 1,\ \ldots,\ m\ .$$

It is well known [11] that when $\tilde{x}_i$ is optimum such numbers (μ_i) exist that

$$\operatorname{grad} F_i(\tilde{x}_i) = \mu_i \operatorname{grad} H_i(\tilde{x}_i),\ i = 1,\ \ldots,\ m\ .$$

Hence, for arbitrary $h_i \in X$, we get also

$$dF_i(\tilde{x}_i,\ h_i) = \mu_i\ dH_i(\tilde{x}_i,\ h_i),\ i = 1,\ \ldots,\ m\ .$$

Then we obtain

$$\frac{d\tilde{f}_i}{d\tilde{y}_i} = \mu_i\ ,\ i = 1,\ \ldots,\ m$$

what completes the proof of (11).

It should be observed that the knowledge of μ_i, $i=1,\ldots, m$ allows to construct the gradient of $\sum_{i=1}^{m} f_i(y_i)$ which enters into the problem of (8), (9).

Then the new admissible values of $\tilde{y}_i$ for the problem (8),

(9) can be determined, which yield greater value of $\sum_{1}^{m} f_i(y_i)$.

These values of $\tilde{y}_i$ can be used for the next solution of (5), (6), (7). Then the following iterational procedure follows (see also Fig. 1):

(i) Starting with an arbitrary admissible value $y_i^{(0)}$, $i = 1, \ldots, m$, the I-level controllers $C_1, \ldots, C_m$, solve the local problems (5), (6), (7) which yield $\mu_i^{(0)}$, $i = 1, \ldots, m$.

(ii) The values $\mu_i^{(0)}$, $i = 1, \ldots, m$, are being sent to the II-level controller C which, by a gradient method, computes the next approximation $y_i^{(1)}$, $i = 1, \ldots, m$, for y_i. These values are being sent to the I-level controllers $C_1, \ldots, C_m$.

Then the process can be repeated yielding an iterational algorithm. When the above formulated iterational process is being derived there is usually such a stage during which the constraint (9) is being violated. When the gradient projection algorithm is being used the new admissible direction should be determined by projection of the gradient on the constraint edges.

The convergence of such a method has all the known properties of gradient methods. For a concrete problem the convergence can be usually investigated in a more complete manner.

The advantage of the described method lies in the fact that the knowledge of the explicit form of $f_i(y_i)$, $i = 1, \ldots, m$, is not necesaary.

4. Extension

In a similar way a problem can be solved in which instead of (6) one gets several local constraints $G_{ij}(x_i) \geq 0, j = 1, \ldots, n$, $i = 1, \ldots, m$, and in the global problem one gets

$$\sum_{i=1}^{m} H_{ij}(x_i) \geq h_j, \quad j = 1, \ldots, p, \tag{14}$$

where h_j - given real numbers.

In a special case when $H_{ij}(x_i) = \alpha_{ij}H_i(x_i)$, $i = 1,\dots, m$, $j = 1, \dots, p$, and α_{ij} are real numbers we get(instead of (9)):

$$\sum_{i=1}^{m} y_i \alpha_{ij} \geqslant h_j \;, \quad j = 1, \dots, p \;. \tag{15}$$

The problem of maximizing (8) , with the constraints (15) can be also solved by gradient technique.

One should observe that the constraints (14),(15) belong to the class of the operator-interactions. There exists also a class of integral constraints such as

$$\sum_{i=1}^{m} \int_0^t k_i(t - \tau)x_i(\tau)d\tau \;\geqslant\; c(t), \; t \in [0,T], \tag{16}$$

where $c(t)$, $k_i(\tau)$ - given functions.

Introducing a discrete version of (16) one gets the system of inequalities

$$\sum_{i=1}^{m} \sum_{\nu=1}^{j} k_i\left[(j - \nu)\,\Delta T\right]x_i(\nu \cdot \Delta T)\,\Delta T \geqslant c(j\cdot \Delta T), \; j = 1,\dots,p$$

where $p = T/\Delta T$, ΔT - a small enough interval. The present form of interaction belongs to the class (15).

Obviously, the described method can be also used for the solution of problems win which it is necessary to maximize

$$\sum_{i=1}^{m} F_i(x_i)$$

subject to

$$G_i(x_i) \leqslant 0, \quad i = 1, \dots, m \,,$$

$$\sum_{i=1}^{m} H_i(x_i) \leqslant h \;,$$

where F_i, G_i, H_i, $i = 1, \dots, m$, are convex and differentiable.

It is also possible to solve problems in which the constrains (6),(14) contain equality constraints or when x_i are multi-variable e. g. - vectors in E^n. However, before that method is used, the problem should be reduced to a standard form of (2) - (4). In certain cases that requires the introduction of auxiliary variables.

It should be also noted that there exist nonlinear transformations which can be used for transformation of the optimalized functionals to the required separable, additive form (2), (4). If for example

$$\phi(x_1, \ldots, x_m) = \prod_{i=1}^{m} \left[\phi_i(x_i)\right]^{\alpha_i},$$

where $\phi_i(x_i)$ and α_i are nonnegative, introducing the functions

$$F_i(x_i) = \ln \phi_i(x_i),\ i = 1, \ldots, m ,$$
$$F(x_1, \ldots, x_m) = \ln \phi(x_1, \ldots, x_m) ,$$

we get

$$F(x_1, \ldots, x_m) = \sum_{i=1}^{m} \alpha_i F_i(x_i)$$

as required by (2).

5. Optimalization of a Thermal-hydro-power System

A typical example of the complex system where the proposed method can be used is an integrated power system, including hydro and thermal power stations.

Assume that the instantaneous cost of generation of electric power P_c in the theraml station $F(P_c)$, where the function $F(P_c)$ is positive, increasing and strongly convex, $F'(0) > 0$. The cost of power generation in the interval $[0,T]$ is equal

$$C = \int_0^T F\left[P_c(t)\right] dt \ . \tag{17}$$

It is assumed also that

$$P_{min} \leq P_c(t) \leq P_{max} , \tag{18}$$

where P_{min}, P_{max} are given positive quantities.

The hydro-station is characterized by the function $P_H[q]$, where P_H is the power generated by the station and q is the rate of water-flow through the hydro-turbine. It is assumed that $P_H[q]$ is positive, increasing and convex, $P_H'(0) > 0$.

Denote by $r(t)$ the rate of water inflow and by $v(t)$ the instantaneous water inventory in the reservoir. One gets an obvious relation

$$v(t) = V_0 - V_{min} + \int_0^t r(\tau)\, d\tau , \tag{19}$$

where: V_0 - amount of water at $t = 0$, V_{min} - minimum admissible water amount in the reservoir.

Consider a power system which consists of hydro and a thermal stations that should generate $P(t)$ units of power for common load where $P(t)$ is a given function. It is necessary to find the best strategy of power $P_c = \overline{P}_c$ and water distribution $q(t) = \overline{q}(t)$ which minimize (17) subject to the constraint (18) and

$$P_H[q(t)] + P_c(t) = P(t) , \tag{20}$$

$$0 \leq q(t) \leq Q , \tag{21}$$

$$\int_0^t q(\tau)d\tau \leq v(t) , \tag{22}$$

where Q is a given positive number.

The above formulated problem has been investigated by many authors [8], using variational calculus, maximum principle and dynamic programming. In one paper a method based on the so called <u>Lagrange functionals</u> has been applied and an explicit solution of the optimalization problems has been given [6]. Further extensions of that method are included in the paper .

In the case of an integrated power system, which consists of n thermal- and m hydro-stations with one reservoir the global cost functional becomes

$$C = \sum_{i=1}^{n} \int_{0}^{T} F_i\left[P_{ci}(t)\right]dt \,, \tag{23}$$

where $F_i\left[P_{ci}\right]$ - given cost functions.

It is required to find such strategies $P_{ci}(t) = \overline{P}_{ci}(t)$, $i = 1, \ldots, n$, and water distribution $q_i(t) = \overline{q}_i(t)$, $i = 1, \ldots, m$, of hydro-stations, characterized by function $P_{Hi}\left[q_i\right]$, which minimize (24) subject to the constraints

$$\sum_{i=1}^{m} P_{Hi}\left[q_i(t)\right] + \sum_{i=1}^{n} P_{ci}(t) = P(t) \,, \tag{24}$$

$$P_{i\ min} \leqslant P_{ci}(t) \leqslant P_{i\ max} \,, \tag{25}$$

$$0 \leqslant q_i(t) \leqslant Q_i \,, \quad i = 1, 2, \ldots, m \tag{26}$$

$$\sum_{i=1}^{m} \int_{0}^{t} q_i(\tau)d\tau \leqslant v(t) \,, \tag{27}$$

where $P_{i\ min}$, $P_{i\ max}$, Q_i - given quantities.

It can be observed that (27) represents the operator-interaction introduced by the common reservoir. One gets a similar type of interaction in the case of several reservoirs when the hydro-stations are situated in tandem connection e. g. on the same river, with distances small enough, so the water delays can be neglected.

In order to solve the optimalization problem by the method of section 3, it is necessary to change it into a discrete version. Dividing $[0, T]$ into p sub-intervals $[0, \Delta T] \ldots$ $\ldots [(p-1)\Delta T, p \cdot \Delta T]$, $p = T/\Delta T$, we get instead of (23)-(27) the problem of minimalization of

$$C^* = \sum_{i=1}^{n} \sum_{\nu=1}^{p} F_i\left[P_{ci}(\nu \cdot \Delta T)\right] \Delta T$$

subject to the constraints:

$$\sum_{i=1}^{m} \left[P_{Hi}\, q_i(\nu \cdot \Delta T)\right] + \sum_{i=1}^{n} P_{ci}(\nu \cdot \Delta T) = P(\nu \cdot \Delta T)$$

$$P_{i\ min} \leqslant P_{ci}(\nu \cdot \Delta T) \leqslant P_{i\ max}$$

$$0 \leqslant q_i(\nu \cdot \Delta T) \leqslant Q_i, \quad i = 1, \ldots, m$$

$$\sum_{i=1}^{m} \sum_{j=1}^{\nu} q_i(j \cdot \Delta T)\, \Delta T \leqslant v(\nu \cdot \Delta T)$$

$$\nu = 1, 2, \ldots, p .$$

In the process of computation, according to the two-level algorithm of section 3, it is expedient to use the dynamic sub-problem solutions (derived, for instance, in papers [6,7]). When ΔT is small enough, the error due to discretization can be neglected.

It should be observed that the two-level optimalization algorithm is very usefull in the case when a decentralized system of energy dispatching centers is being used.

References

1. C.B. Brosilov, L.S. Lasdon, I.D. Pearson, D. Macko, Y. Takahara: Papers on multi-level control systems. Case Institute of Technology, Systems Research Center, SCR-70-A-65-25 1965.
2. R. Kulikowski: Optimum control of aggregated multi-level systems. Proc. III IFAC Congr., London 1966.
3. R. Kulikowski: Optimum control of multi-dimensional and multi-level systems. In: Advances in Control Theory. Ed. by C. Leondes. Vol. 4. New York 1966 Academic Press Inc.
4. R. Kulikowski: Optimum control of dynamic interacting systems. Arch. Automat. i Telemechan. 1966 z. 2.
5. R. Kulikowski: Optimization of aggregated dynamic systems. Arch. Automat. i Telemechan. 1966 z. 3.
6. R. Kulikowski: On optimum control of nonlinear, dynamic industrial processes. Arch. Automat. i Telemechan. 1967 z. 1.

77. J. Majerczyk-Gómułka, K. Makowski: Wyznaczanie optymalnego sterowania procesami dynamicznymi metodą funkcjonałów Lagrange'a. Warszawa 1967. Prace IPPT nr 9, in polish.

8. L.K. Kirchmayer, R.I. Ringle: Optimal control of thermal-hydro system operation. Proc. II IFAC Congr., Basel 1963.

9. J.B. Rosen: The gradient projection method of nonlinear programming. J. Ind. Appl. Math. 1960 p. 181-217, 1961 p. 514-532.

10. I.L. Sanders: A nonlinear decomposition principle. Operation Research 1965 No. 2.

11. M.M. Weinberg: Variational methods of investigation of nonlinear operators. Moscow 1956, in Russian.

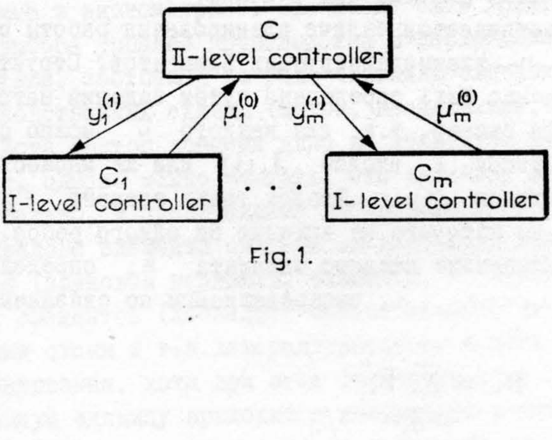

Fig. 1.

ПРИНЦИП ДЕЦЕНТРАЛИЗАЦИИ ПРИ ОПТИМИЗАЦИИ СЛОЖНЫХ СИСТЕМ

Первозванский А.А.
профессор, Политехнический институт
Ленинград, СССР

Работа состоит из двух частей. В первой из них анализируются возможности разложения задачи оптимального планирования для системы достаточно общей структуры на ряд экстремальных задач для отдельных элементов. Выясняются особенности влияния взаимных поставок на эффективность работы каждого элемента и предлагается простейшая локальная аппроксимация для таких зависимостей. На этой основе во второй части анализируется влияние неравномерности поставок внутри планового периода на среднюю эффективность работы элемента с учётом возможности оптимизации политики резервирования, что приводит к рекомендациям о необходимости коррекции постановки задачи планирования для системы в целом.

I. Рассматривается задача планирования работы системы, состоящей из n взаимодействующих элементов. Структура взаимодействия может быть определена путём задания матрицы инциденций графа связей, т.е. для каждого i можно определить множество индексов j входов $\mathcal{J}_1(i)$ или же множество индексов j выходов $\mathcal{J}_2(i)$. Введём также элемент ε_{n+1} (вершину графа), из которого не выходит ни одного ребра.

Функционирование каждого элемента ε_i определяется выпуском P_i, $i=1,2,\ldots,n$, распределяемым по связанным с ним элементам

$$P_i = \sum_{j\in \mathcal{J}_2(i)} P''_{ij} \tag{I.I}$$

причём уровень выпуска может изменяться в пределах, допустимых ограничениями

$$\{ u_i, P_i, P'_{ij} \} \in R_i, \tag{I.2}$$

где u_i - вектор, характеризующий интенсивность процессов в самом элементе, P'_{ij}, $j \in \mathcal{J}_1(i)$,- вектора, характеризующие потребляемую им долю выпуска других элементов, R_i - некоторые замкнутые области.

Отметим, что иногда удобно разделять ограничения на две группы: системные ограничения, в которые входят уровни потребления продукции других элементов, и собственные (местные), характеризующие возможности только данного элемента. Предполагается, что функционирование системы в целом подчинено экстремальному принципу, выраженному в аддитивной форме

$$F = \sum_{i \in J_i(n+1)} F_i \left(P''_{i,n+1} \right) \to \max \tag{I.3}$$

и условиям баланса на связях

$$P'_{ij} = P''_{ij} \quad , \quad i,j = 1,2,\ldots,n . \tag{I.4}$$

Решение сформулированной экстремальной задачи и будет именоваться решением задачи оптимального планирования для системы в детерминированной постановке. По своей формулировке она является задачей математического программирования и структурно обобщает классические схемы Л.В.Канторовича и Д.Гейла, используемые в эконометрике. Она может быть трактована как проблема выбора уровней производства и распределения продукции в системе, состоящей из ряда частично связанных между собой производственных единиц (цехов, предприятий, объединений и т.п.). Хотя фактор времени явно не участвует в формировании модели, однако можно считать, что заданные ограничения, а следовательно, и производимый выбор должны определять поведение каждого элемента системы на некотором конечном отрезке времени (плановом периоде), возможно, не идентичном для различных элементов (последнее обстоятельство позволяет ввести в рамки схемы и т.н.детерминированные модели перспективного планирования, хотя при этом физически одну и ту же производственную единицу приходится изображать различными элементами схемы, соответствующими ее состоянию на последовательных подпериодах планирования).

После формирования проблемы планирования она может, по крайней мере, в принципе решаться любым известным методом математического программирования. Однако особенности модели как большой системы, налагают существенные ограничения на выбор метода. Эти ограничения двух типов: а) информационные, связанные с трудностями сбора и хранения всех данных, с помощью которых осуществляется конкретное описание всех элементов системы, в одном месте (центре); б) вычислительные, вызванные практической неспособностью современной вычислительной техники оперировать с задачами большого объема.

Поэтому в данной работе внимание уделяется только методам, в которых используется принцип децентрализации. Под этим принципом понимается следующее: основная задача должна быть расчленена на n экстремальных подзадач, независимых в том смысле, что для их взаимного согласования не требуется прямой информации о характере ограничений для каждого из элементов системы.

При более жесткой трактовке этого принципа выдвигается дополнительное требование о том, что согласование подзадач может производиться не в единственном центре, а в нескольких, каждый из которых обладает лишь частичной информацией.

Математическая трактовка идеи децентрализации (хотя и в ограниченной форме) возникла на базе работ Эрроу, Гурвица, Удзавы по градиентным схемам поиска экстремума [I,2], далее получила новый толчок при создании метода декомпозиции Данцига-Вулфа [3], и в настоящее время широко обсуждается, главным образом, в связи с различными схемами т.н."блочного" программирования [4 ÷ 8] и др.

Можно наметить (опуская ряд существенных математических деталей) два различных подхода к формированию подзадач.

Первый из них носит общий характер и основан на использовании основной теоремы Куна и Таккера об эквивалентности (при известных ограничениях) исходной задачи и проблемы разыскания седловой точки функции Лагранжа.

$$\varphi = \sum_{i \in \mathcal{J}_i(n+1)} F_i\left(P''_{i,n+1}\right) + \sum_i \sum_j \lambda_{ij}\left(P''_{ij} - P'_{ij}\right) \qquad (I.5)$$

при условиях (I,I ÷ I.2).
Тогда очевидно распадение на n подзадач для элементов

$$f_i = \max\left[F_i\left(P''_{i,n+1}\right) + \sum_{j \in \mathcal{J}_2(i)} \lambda_{ij} P''_{ij} - \sum_{j \in \mathcal{J}_1(i)} \lambda_{ji} P'_{ji}\right],$$

$$P_i = \sum_{j \in \mathcal{J}_2(i)} P''_{ij}, \quad \{u_i, P_i, P'_{ij}\} \in R_i, \qquad (I.6)$$

при решении которых "цены" λ_{ij} рассматриваются как параметры, и одну задачу для "центра", являющуюся задачей оптимального согласования "цен"

$$\varphi_0 = \min \left\{ \varphi = \sum_{i=1}^{n} f_i [\lambda_{ij}, i \in J_1(i); j \in J_2(i)] \right\} \qquad (I.7)$$

Эта схема может условно именоваться схемой "купли-продажи" промежуточной продукции, причём подзадачи трактуются как задачи максимизации прибыли для каждого элемента в ценах на промежуточную продукцию, согласуемых "центром".

Второй подход предполагает, что каждый из элементов системы участвует в формировании конечной продукции, т.е.множество $J_1(n+1)$ включает все $i = 1, 2, \ldots, n$.

Тогда, рассматривая в качестве параметров уровни распределяемой продукции, приходим к n подзадачам вида

$$\mathcal{F}_i = \max \left[F_i(P''_{i,n+1}) / \sum_{j \in J_2(i)} P''_{ij} = P_i ; \{u_i, P_i, P'_{ij}\} \in R_i \right] \qquad (I.8)$$

и одной задаче для "центра", являющейся задачей оптимального балансирования промежуточной продукции:

$$F = \max \left\{ \mathcal{F} = \sum_i \mathcal{F}_i [P'_{ij}, j \in J_1(i); P''_{ij}, j \in J_2(i)] \right\}, \qquad (I.9)$$

$$P'_{ij} = P''_{ij}, \quad i, j = 1, 2, \ldots, n.$$

Такая схема может быть условно названа схемой распределения ресурсов или схемой "оптимального баланса". При этом внутри каждого элемента производится оптимизация по выпуску конечной продукции при предписанных "центром" уровнях взаимных поставок.

Описанные разьиения являются чисто формальными. Действительно, не существует другого способа указать "цены" (объективно обусловленные оценки, по терминологии Л.В.Канторовича) или уровни промежуточной продукции, при которых осуществляется оптимальное согласование подзадач помимо построения решения задачи в целом. Однако возможность использования итеративных процедур, т.е. процедур постепенного согласования, делает этот подход достаточно эффективным.

Почти все известные процедуры являются вариантами обобщенного градиентного спуска или метода возможных направлений, применяемого к задаче центра. Основная идея применительно к (I.7) сводится к следующему: "центром" выбирается некоторый произвольный набор "цен" λ^{i}_{ij} (в реальных задачах этот выбор, естественно, диктуется априорными практическими соображениями), этот выбор сообщается каждому элементу, внутри которого производится решение задачи типа (I.6), приводящее к оптимальным при заданных λ_{ij} значениях уровня производства u_i, поставок P''_{ij}, $j \in J_2(i)$, и внешнего (для данного элемента) потребления P'_{ij}, $j \in J_1(i)$. Кроме того, выявляется тенденция изменения функции цели при малых отклонениях "цен" λ_{ij} от уровня λ^0_{ij}, иначе говоря, строится локальная аппроксимация $f_i(\lambda_{ij})$ в некоторой малой окрестности λ^0_{ij}.

Именно эта локальная аппроксимация сообщается в "центр", где тем самым выявляется тенденция изменения функции φ в целом, что позволяет найти направление изменения системы "цен", на котором значение этой функции быстрейшим образом уменьшается.

После осуществления шага по выбранному направлению, приводящего к нахождению новых значений $\lambda_{ij} = \lambda^{(1)}_{ij}$, процедура, естественно, может быть повторена.

Выбор способа локальной аппроксимации и величины шага по направлению убывания является специфическим в каждом из существующих методов.

Отметим лишь некоторые общие особенности, возникающие в том случае, когда исходная задача, а следовательно - и подзадачи формализуются в виде задач линейного программирования.

Здесь локальная аппроксимация поведения функций f_i, точно совпадающая с истинным поведением в некоторой конечной окрестности λ^0_{ij} имеет вид

$$\hat{f}_i = f_{ij}(\lambda^0_{ij}) + \min_{s \in S^0_i} \Big\{ \sum_{j \in J_2(i)} \Delta\lambda_{ij} P''_{ij,s} - \sum_{j \in J_1(i)} \Delta\lambda_{ij} P'_{ij,s} \Big\}, \qquad (I.10)$$

где S^0_i - множество оптимальных базисов задачи (I.7) при $\lambda_{ij} = \lambda^0_{ij}$, а $P'_{ij,s}$, $P''_{ij,s}$ - соответствующих этим базисам оптимальных значений переменных P'_{ij}, P''_{ij}. Отсюда,в частности, ясно, что функции $f_i(\lambda_{ij})$, вообще говоря, недифференцируемы. Если же при $\lambda_{ij} = \lambda^0_{ij}$ решение (I.7) единственно, то градиент $f_i(\lambda_{ij})$ в этой точке существует и находится сразу, как только известно это оптимальное решение.

Вычисление же градиента функции φ в целом, если все задачи типа (I.7) имеют единственное решение, сводится к подсчёту небалансов на связях:

$$\operatorname{grad}_{\lambda_{ij}} \varphi \Big|_{\lambda_{ij}=\lambda_{ij}^{0}} = P''_{ij,c} - P'_{ij,c} \qquad (I.II)$$

Указанное обстоятельство существенно облегчает организацию градиентного спуска. Более того, отсюда вытекает принципиально важное обстоятельство: определение направления изменения "цен" может вестись без участия единого центра, а лишь путем согласования результатов планирования непосредственно связанных элементов. Вместе с тем очевидно, что условие однозначности выполняется не всегда и заведомо нарушается в точке экстремума функции φ, следствием чего является, в частности, отсутствие строгой сходимости градиентного спуска при постоянных коэффициентах пропорциональности [I]. Однако ряд приемов, например, уменьшение коэффициентов с ростом номера итерации [9], позволяет избежать указанных неприятностей. Более того, при решении практических задач требование точной сходимости не столь существенно, скорее важно повышение быстроты движения, что может быть обеспечено, например, использованием конечно-шаговых схем с применением упрощенных способов локальной аппроксимации [I0].

Отметим далее, что все сказанное в полной мере относится и к второму способу применения принципа децентрализации. Здесь также применимы для решения задачи "центра" схемы обобщенного градиентного спуска, но уже в пространстве переменных P'_{ij}, P''_{ij}. Функции $\mathfrak{F}_i$ также являются кусочно-линейными функциями, если исходная задача является задачей линейного программирования. Их локальные аппроксимации могут быть построены с помощью оптимальных базисных решений задач, двойственных к (I.8). Если таковые решения единственны, т.е. задачи (I.8) при заданных значениях параметров невырожденны, то возможно определение градиентов функций $\mathfrak{F}_i$ с помощью оптимальных двойственных переменных (объективно обусловленных оценок). В окрестности оптимума допустима лишь кусочно-линейная локально-точная аппроксимация вида

$$\hat{\mathfrak{F}}_i = \mathfrak{F}_i\left[P'_{ij,opt};\, P''_{ij,opt}\right] + \min_{\ell \in \ell_j^{0}} \Big\{ \sum_j \lambda'_{ij,\ell} \cdot \Delta P'_{ij} - \sum_j \lambda''_{ij} \cdot \Delta P''_{ij} \Big\}, \qquad (I.I2)$$

где ℓ_j^c - множество, оптимальных базисов двойственной задачи, а $\lambda'_{ij,\ell}$, $\lambda''_{ij,\ell}$ - соответствующие значения двойственных переменных. Проведение итеративных процедур осуществимо почти аналогично вышеописанному для первого способа разбиения, однако оно производится при точном соблюдении условий баланса на связях в каждой итерации, что хотя и осложняется ее реализацию, но повышает ценность построенных приближений, каждое из которых представляет собой субоптимальный план задачи в целом.

Отметим, что результатом установления плана (программы) обычно считается задание для каждого элемента системы уровня конечной продукции и взаимных поставок интегрально на плановом периоде. Однако в силу неизбежной флюктуативности условий точная реализация плана оказывается невозможной, поэтому при осуществлении управления, т.е. реализации плана, естественно возникает вопрос о компенсации отклонений от программы. При этом,ввиду множественности показателей плана, существенным является вопрос о соизмерении важности отклонений по различным показателям, т.е. степени влияния этих отклонений на эффективность работы системы. В практике управления это соизмерение производится субъективно. Вместе с тем осуществление планирования по принципу децентрализации, и в этом его главное достоинство, позволяет при построении плана не только найти оптимальные показатели, но и произвести объективное соизмерение эффективности малых отклонений от них, поскольку в ходе построения плана строится и локальная аппроксимация зависимостей местных целевых функций от внешних параметров. Существенно, что такая аппроксимация даже в малой окрестности оптимального плана является нелинейной, а это предопределяет невозможность использования таких "линейных" показателей, как суммарная стоимость отклонений, вычисленная в ценах, не зависящих от уровня или направления отклонений.

2. Перейдем далее к анализу влияния некоторых флюктуативных факторов на эффективность реализации программы в ходе ее осуществления и обратного влияния этих факторов на схему построения самого плана.

Очевидно, что задание программы взаимных поставок интегрально за плановый период не обусловливает распределения этих поставок в течение этого периода.

Этот недетерминизм, присущий самому способу планирования, может приводить (и практически приводит) к возникновению неравномерности в уровне поставок.

Выделим некоторый элемент системы и изучим изменение его конечной продукции при отклонениях уровня поставок этому элементу от остальных от равномерного.

Разобьем плановый период на подпериоды, перенумеровав их индексом κ, $\kappa = 1, 2, \ldots, \mathcal{K}$. Предположим, что собственные ресурсы этого элемента остаются неизменными в течение всего периода и что его функционирование подчинено сформулированному выше локальному экстремальному принципу, согласованному (на оптимальном в целом за период плане взаимных поставок) с экстремальным принципом для системы в целом. Тогда за каждый подпериод изменение эффективности по выпуску конечной продукции при выполнении требований по поставкам из данного элемента при малых отклонениях поставок в него от заданного уровня определяется формулой, аналогичной (I.I2). Эта кусочно-линейная зависимость может быть эффективно приближена с помощью системы коэффициентов (маргинальных значений задачи по [II]), характеризующих изменение функции $\mathfrak{F}_i$ при возрастании или убывании каждой из компонент поставляемой продукции от оптимальной программы по отдельности.

Опуская индексы, характеризующие номер элемента и его поставщиков, эту локально аппроксимирующую зависимость можно записать в виде [10] :

$$\bar{\bar{\mathfrak{F}}}_\kappa = \mathfrak{F}_\kappa(P_{\kappa,opt}) + \sum_{z} \min(\lambda_z^{+}\Delta P_{z,\kappa};\, \lambda_z^{-}\Delta P_{z,\kappa}) = \mathfrak{F}_{\kappa,opt} + \sum_{z} \Delta\mathfrak{F}_{z,\kappa}, \qquad (2.I)$$

где λ_z^{+}, λ_z^{-} – соответственно правые и левые частные производные целевой функции по z -му ресурсу. Можно показать, что $\bar{\bar{\mathfrak{F}}} \leq \hat{\mathfrak{F}}$.

Преимуществом аппроксимации (2.I) является возможность рассматривать влияние изменений по каждой компоненте отдельно, различие между компонентами проявляется лишь в различии маргинальных значений.

Поэтому в дальнейшем будем опускать и индекс компоненты ресурса $P_{z,\kappa}$ и соответствующего ему изменения $\Delta\mathfrak{F}_{z,\kappa}$ функции цели и примем обозначения:

$$\Delta p_{z,k} \equiv v_k ; \Delta \mathfrak{F}_{t,k} \equiv \psi(v_k) = \begin{cases} \lambda^+ v_k , & v_k \geq 0, \\ \lambda^- v_k , & v_k \leq 0. \end{cases} \qquad (2.2)$$

Отметим также, что всегда $\lambda^- \geq \lambda^+ \geq 0$.
Величины v_k, $k=1,2,\ldots,\mathcal{K}$, характеризуют отклонение потребляемого данным элементом системы количества рассматриваемой компоненты ресурса от планового задания, представляющей собой $1/\mathcal{K}$ часть плановых поставок за период в целом.

Можно также ввести величины η_k, характеризующие отклонения реальных поставок за подпериод k от указанного планового уровня. Если поставки не могут накапливаться, то $v_{k+1} = \eta_k$. Если же возможно накопление излишков ресурсов, то выбор величины v_k ограничен наличием на складе к началу подпериода

$$v_k \leq y_k , \qquad (2.3)$$

причем

$$y_{k+1} = y_k + \eta_k - v_k , \quad k = 1, 2, \ldots, \mathcal{K}. \qquad (2.4)$$

Таким образом, общая проблема учета неравномерности поставок сведена к классической одномерной задаче управления запасами, частный случай которой - управление запасами воды в водохранилище при совместной работе тепловых и гидроэлектрических станций был рассмотрен ещё Карлиным и Скарфом в [I2]. Проведенный общесистемный анализ и использование упрощенной локальной аппроксимации (2.I) придают этой проблеме достаточно общий характер.

Опишем конкретные результаты анализа при условии использования оптимальной многошаговой политики на всем периоде и простейшей политики $v_{k+1} = \eta_k$. При этом предположим, что η_k равной вероятностью отклоняется от нуля (планового уровня) на величину $\pm a$, будучи в среднем равной нулю. Тогда, если обозначить через $S_t(y)$ величину средних потерь, вызванных неравномерностью за t подпериодов до конца планового периода, при начальном запасе y и оптимальной политике расходования, то

$$S_t(y) = \min_{v \leq y} \left[-\psi(v) + \int S_{t-1}(y+\eta-v)\, W(\eta)\, d\eta \right] =$$

$$= \min_{v \leq y} \left[-\psi(v) + \tfrac{1}{2} S_{t-1}(y+a-v) + \tfrac{1}{2} S_{t-1}(y-a-v) \right] \qquad (2.5)$$

$$S_1(y) = \begin{cases} -\lambda^+ y, & y \geq 0 \\ -\lambda^- y, & y \leq 0 \end{cases}$$

Можно показать, что оптимальной является политика:

$$v_t = \begin{cases} 0, & y \geq 0; \\ y, & y \leq 0, \end{cases} \quad t = 2, \ldots, \mathcal{K}; \quad v_1 = y. \tag{2.6}$$

Тогда

$$S_t(y) = \begin{cases} \frac{1}{2} S_{t-1}(y+a) + \frac{1}{2} S_{t-1}(y-a), & y \geq 0, \\ -\lambda^- y + \frac{1}{2} S_{t-1}(a) + \frac{1}{2} S_{t-1}(-a); & y \leq 0 \end{cases} \tag{2.7}$$

$$S_1(y) = \begin{cases} -\lambda^+ y; & y > 0 \\ -\lambda^- y, & y < 0. \end{cases}$$

При использовании простейшей политики

$$\bar{S}_t(y) = -\psi(y) + \frac{1}{2}[\bar{S}_{t-1}(a) + \bar{S}_{t-1}(-a)] = -\psi(y) + (t-1)\frac{\lambda^- - \lambda^+}{2} a \tag{2.8}$$

Графики функций $S_t(y)$ при $t = 1, 2, \ldots, 6$, приведены на рис.I.

Из этого рассмотрения можно сделать следующие выводы:

I. Флюктуации поставок, равные нулю в среднем , приводят к возникновению средних потерь, вообще говоря, отличных от нуля.

2. Потери при простейшей политике и отсутствии начального резерва равны

$$(\mathcal{K} - 1)\frac{\lambda^- - \lambda^+}{2} a,$$

т.е. растут пропорционально числу подпериодов.

3. При наличии резерва и оптимальном использовании накоплений можно добиться снижения средних потерь. Даже при отсутствии резерва средние потери при $\mathcal{K} = 6$ снижаются приблизительно вдвое по сравнению с результатом простейшей политики.

Можно также показать, что вероятность полного отсутствия потерь резко зависит от величины начального резерва y_0 и равна $2^{N-\mathcal{K}}$, если $N = y_0/a < \mathcal{K}$.
Описанный эффект влияния перавномерности поставок на эффективность системы является существенным, но отражает лишь одну сторону дела. Действительно, в ходе диализа нами неявно предполагалось, что задача оптимизации для выделенного

элемента системы разрешима не только при плановом уровне поставок, но и при их флюктуациях. Иначе говоря, что и при флюктуациях удается выполнять ограничения по плановым уровням поставок из выделенного элемента. Вообще же говоря, это не имеет место: при отклонениях поставок часто оказывается невозможным выполнить требования по внутрисистемным заказам, а следовательно - при децентрализованном текущем планировании возникает необходимость оптимально соизмерить отклонения от требуемых уровней.

Использование местных функционалов в форме (I.I2) не дает решения этого вопроса, так как сами функции $\mathcal{F}_i(P)$ естественно определены только на области существования решений.

Особенно острой является эта проблема для структуры типа технологической цепи [7,I3] , где элементы объединяются последовательно и так, что лишь последний элемент цепи выдает конечную продукцию с известными оценками компонент.

Вместе с тем, для анализа такой структуры может быть применена схема децентрализации, построенная на базе динамического программирования.

При этом, вообще говоря, необходимо лишь построение зависимости эффективности каждого элемента от располагаемых ресурсов при малом изменении их от оптимального планового уровня. В работе [I3] было показано, что локально-точная аппроксимация функции цели для каждого элемента с номером p технологической цепи, согласованная с известными оценками конечной продукции, может быть представлена в виде кусочно-линейной функции

$$\hat{f}_p(v) = f_{p.opt} + \min_{\ell \in \ell_p^0} \lambda_p v, \tag{2.9}$$

где v - матрица-столбец отклонений от оптимума уровня ресурсов, поставляемых элементу p от предшествующего по технологической цепи. Аналогичный характер имеют и закономерности для зависимостей от v каждого из видов продукции, выпускаемой данным элементом. При переходе к упрощенной аппроксимации типа (2.I), получим, что зависимость отклонения от оптимального плана любой компоненты продукции элемента p от малых отклонений любой компоненты потребляемого из $p-1$ ресурса представима функцией типа $\varphi(v)$ (см.

(2.2)).

Это обстоятельство приводит к эффекту накопления влияния флюктуаций, вызванных неравномерностью взаимных поставок, вдоль технологической цепи, если в ней не предусмотрено наличия необходимого уровня начального резерва (задела) для каждого элемента. Действительно, флюктуации поставок на первый элемент цепи, даже если они в среднем соответствуют плановым, приводят к невозможности выдержать средний уровень поставок от первого элемента ко второму и т.д. В силу этого, действительные средние уровни поставок для последующих элементов могут оказаться столь далекими от плановых, что сами оценки эффективности отклонений, справедливые лишь в окрестности плана, станут неверными. Поэтому само оптимальное планирование может быть эффективным только, если оно учитывает наличие и необходимость затрат на формирование резервов, позволяющих локализовать влияние неравномерности.

Естественно, что создание и сохранение резервов представляется неэкономичным с точки зрения достижения цели системы, однако также оно необходимо и имеет тот же смысл, что введение избыточности в технические системы, составленные из не вполне надежных элементов. Следует подчеркнуть, что здесь идет речь не об эффекте флюктуаций внешних условий,по отношению к которому необходимость резервирования общепринята, а об эффекте внутрисистемных флюктуаций, стимулируемых неполным детерминизмом общесистемного планирования.

Литература

1. Arrow K.J., Hurwicz L., Vzawa H., Studies in linear and nonlinear programming, Stanf., Calif., 1958.
2. Marschak T., Centralization and decentralization in economic organizations, Econometrica, v. 27, 3 (July 1959).
3. Dantzig G.B., Linear programming and extensions, Princet.,N.-Y., 1963.
4. Гольштейн Е.Г., Юдин Д.Б., Новые направления в линейном программировании, "Сов.Радио", М.,1966.
5. Волконский В.А., Оптимальное планирование в условиях большой размерности, экономика и математические методы, 1965, т.I, № 2.

6. Корнаи Н.; Липтак Т., Планирование на двух уровнях, в сб."Применение математики в экономических исследованиях", т.3, "Мысль", 1965.
7. Первозванская Т.Н., Первозванский А.А., Распределение ресурсов между многими предприятиями, экономика и математические методы, 1966, т.2, № 5.
8. Каценеленбойген А.Н., Овсиенко Ю.В., Фраерман Е.Ю., Методические вопросы оптимального планирования в социалистической экономике, ЦЭМИ, АиСССР, М.,1966.
9. Шор Н.З. Принцип обобщенного градиентного спуска в блочном программировании, Кибернетика, № 3, 1967.
10. Первозванская Т.Н., Первозванский А.А., Алгоритм поиска оптимального распределения централизованных ресурсов, Изв.АН СССР, Техническая кибернетика, 1966, № 3.
11. Mills H.D., Marginal values of matrix games and linear programs, "Linear in equalities and related systems", Prince N.-Y., 1956.
12. Arrow K.Y., Karlin S., Scraf H., Studies in the Theory of Inventory and Production, Stanf., Calif., 1958.
13. Первозванская Т.Н., Метод приближения в пространстве целевых функций при решении задач на "узкие места", ЖВМ и МФ, т.7, № 3, 1967.

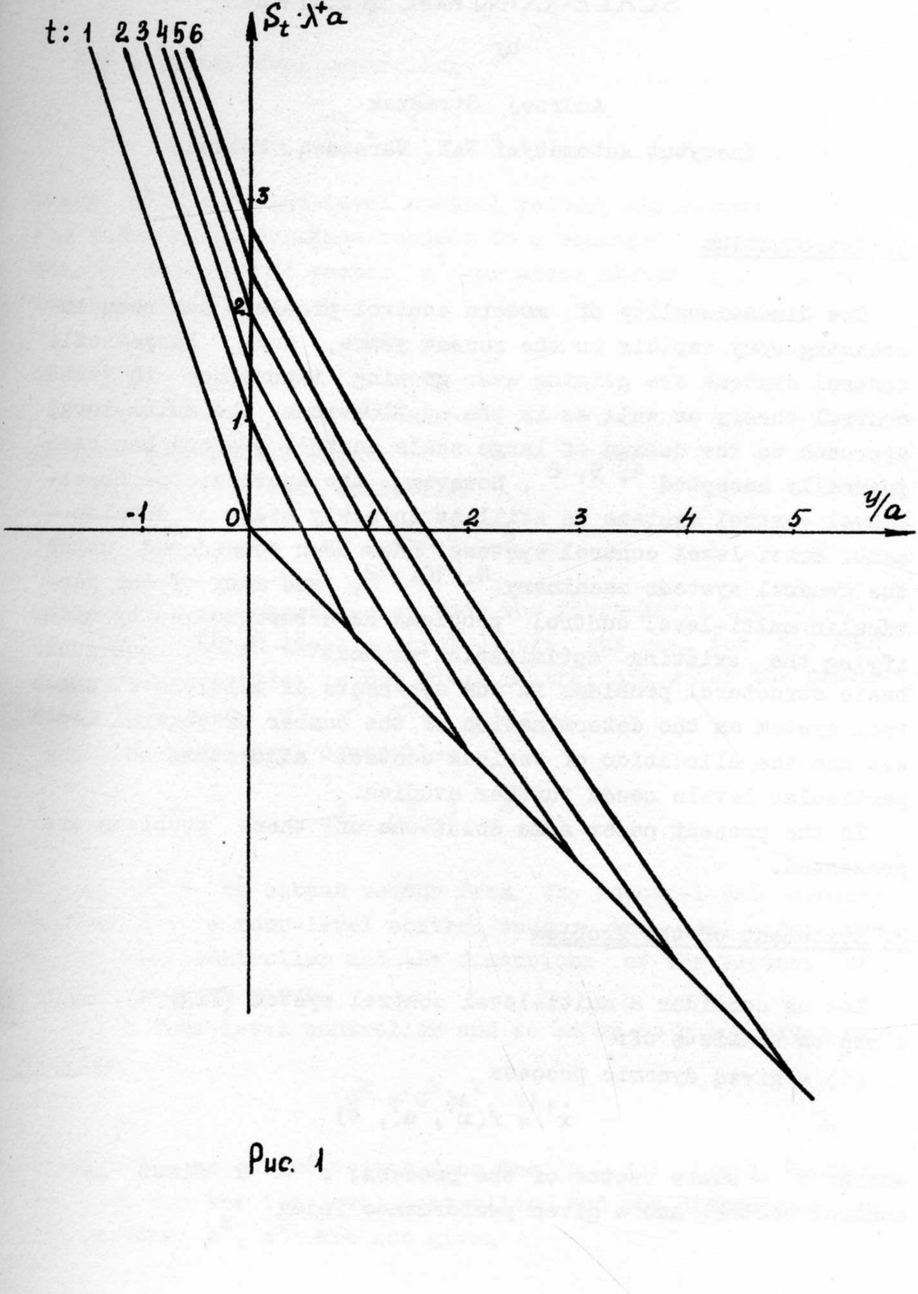

Рис. 1

ON THE SYNTHESIS OF MULTI LEVEL LARGE SCALE CONTROL SYSTEMS

by

Andrzej Straszak

Instytut Automatyki PAN, Warszawa, Poland

1. Introduction

The dimensionality of modern control problems has been increasing very rapidly in the recent years, and large-scale control systems are gaining ever growing importance in the control theory as well as in the applications. The multi-level approach to the design of large scale control systems has been generally accepted [4, 5, 8], however, the synthesis of multi-level control systems is still at an early stage of development. Multi-level control systems have been considered using the general systems machinery [8, 10, 12], and many of the particular multi-level control problems have been solved by modifying the existing optimization methods [5, 9, 11], but such basic structural problems in the synthesis of multi-level control system as the determination of the number of control levels and the allocation of various control algorithms to the particular levels needs further studies.

In the present paper some solutions of these problems are presented.

2. Statement of the Problem

Let us consider a multi-level control system (Fig. 1). Such a system consists of:

(1) a given dynamic process

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

where: x^1 - state vector of the process, u^1 - a first level control vector, and a given performance index

$$I = \int_{t_0}^{t_1} \varphi(x^1, u^1)dt$$

(2) a first-level controller

$$u^1 = C^1(x^1, u^2, t)$$

where u^2 - a second-level control vector, which must minimize the index of performance subject to a boundary condition and subject to (control vector u^1 or state vector x^1) process constraints

$$R(u^1) \leqslant R_0$$

or/and

$$W(x^1) \leqslant W_0$$

if such constraints exist,

(3) a second-level controller

$$u^2 = C^2(x^2, u^3, t)$$

where: x^2 - an output vector from the first-level control system, u^3 - a third-level control vector, the goal for the second-level controller is not given,neither are given the dimensions of vectors u^2, u^3, x^2,

(4) a third-level controller

$$u^3 = C^3(x^3, u^4, t)$$

where: x^3 - an output vector from the second-level control system, u^4 - a four-level control vector, the goal for the third-level controller and the dimensions of the vectors u^3, u^4, x^3 are not given,

(5) a four-level controller and so on up to a top level controller

$$u^k = C^k(x^k, t)$$

where x^k - an output vector from the (k - 1) level control the goal for the top level controller and the dimensions of the vectors u^k, x^k are not given.

It is easy to show that if such a first-level optimal controller exists

$$u_0^1 = C_0^1(x^1,\ u^2,\ t)$$

where $u_0^2 \neq 0$, which minimizes a performance index I, then the first level optimal controller also exists

$$u_0^{1*} = C_0^{1*}(x^1,\ u^2,\ t) = C_0^{1*}(x^1,\ t)$$

where $u_0^2 \equiv 0$, which minimizes the performance index I, and

$$u_0^1 \equiv u_0^{1*}$$

It follows therefrom that if no additional constraints are given, the first level control system is optimal and no higher levels of control are needed. However, if we consider not only process constraints but also controller constraints, for example due to a given bounded date processing ability, then a multi-level control system becomes a necessity.

Let us introduce a complexity measure of a control algorithm [4, 11, 12, 14]. Each control algorithm needs a list of data processing operations: m_α operations α , m_β operations β , m_γ operations γ , etc., hence the complexity measure of a given control algorithm may be introduced in the following form:

$$Z = m_\alpha k_\alpha + m_\beta k_\beta + m_\gamma k_\gamma + \ldots$$

where: k_α - complexity of the operation α , k_β - complexity of the operation β , k_γ - complexity of the operation γ etc.

In general, for a large scale control system it is necessary to assume that the overall list of available data processing operations is limited, hence the overall complexity of the control algorithms must be limited too.

Since

$$Z_g = Z_1 + Z_2 + \ldots + Z_k \leqslant Z_o$$

where: Z_g - the overall complexity of control algorithms in the multi-level system, Z_1 - the complexity of a control algorithm in the first-level control, Z_2 - the complexity of a control algorithm in the second-level control, Z_k -the complexity of a control alforithm in the top-level control, Z_0 - the given complexity of the available data processing operations, and since the complexities of each control algorithm depends on the dimension of control and the type of the control algorithm [11, 12, 14] therefore the synthesis problem for a multi--level control system may be formulated as follows:

Given a set of all the known control algorithms for the first-level and higher-level (supervisory) controls,given the overall available data processing operations, we find an optimal allocation of control algorithm (including determination of the number of control levels,dimensions of the control vector in each control level and types of control) and adjoin the data processing operations such that the optimization of parameters (by adjusting the parameters to given control algorithms) can guide to the most desirable value of the performance index I, and render the overall complexity of control admissible.

3. Control Algorithms in Multi-level Control Systems and their Complexities

In multi-level control systems we have two different classes of control problems. One which is connected with the direct control of the multi-variable process controlled or set of processes controlled by a set of direct controllers, and the other one which is connected with the supervisory control (adjustment) of the set of control systems by supervisory controllers. All direct controllers belong to the first-level control and all supervisory controllers belong to the second- and higher-level of control.

Let us first consider a direct control problem.Since a full description of the process as well as the performance index and all the process constraints are given, we may use in many cases the existing optimal control theory methods [1,3] and try to find the general form of the solution (without determining the values of all parameters precisely)

$$u_0^1 = C_0^1(x^1, t)$$

For example, if

$$\dot{x}^1 = Fx^1 + Gu^1$$

where F, G are $(n^1 \times n^1)$ constant matrices, and

$$I = \int_0^{\infty} \left[(x^1)^T Qx^1 + (u^1)^T Ru^1\right] dt$$

then
$$u_0^1 = C_0^1 x^1$$

where C_0^1 is a $(n^1 \times n^1)$ constant matrix, if

$$I = \int_0^t dt \quad \text{and} \quad \sup |u_i^1| \leqslant l_{oi}$$

then

$$u_{oi}^1 = l_{oi} \text{ sign } h_i(x_0^1)$$

where $h(x_0^1)$ - a nonlinear function.

In spite of the optimal solutions in the control system practice and literature, many sub-optimal solutions are known which require a less complex list of data processing operations, for example, more linear operations instead of nonlinear ones.

If general forms of solutions are found, we can determine lists of data processing operations for each optimal and all sub-optimal control algorithms.

For example, for linear multi-variable direct control algorithms we require:

$(n^1)^2$ operations of multiplication by a constant, and
n^1 operations of summation

therefore $Z = (n^1)^2 k_\alpha + (n^1) k_\beta$, where n^1 - dimension of the control problem; for a non-linear multi-variable direct control algorithm

$$Z = Z\left[n^1, (n^1)^2, (n^1)^3, \ldots\right]$$

Since in multi-level control systems we have a set of direct controllers instead of one, we have to decompose the direct control problem into a set of direct control problems.

By decomposition of the overall direct control problem we obtain the following sub-problems:

Given

$$\dot{x}_i^1 = f_i(x_i^1, u_i^1, \xi_i, t)$$

$$R(u_i^1) \leqslant \rho_i(R_0, u_j^1), \quad u^1 = (u_i^1, u_j^1)$$

$$W(x_i^1) \leqslant \omega_i(W_0, x_j^1), \quad \xi_i = \Phi_i(x_j^1, u_j^1), \; I_i = \int_{t_0}^{t_1} \varphi_i(x_i^1, u_i^1)dt$$

$$x^1 = (x_i^1, x_j^1)$$

find

$$u_i^1 = C_i^1(x_i^1, \xi_i, \rho_i, \omega_i, t)$$

then find the list of data processing operations, and

$$Z\left[C_i^1(x_i^1, \xi_i, \rho_i, \omega_i, t)\right] = Z(n_i^1)$$

where $n_i^1 = \dim u_i^1$.

Some solutions of this type of control problems are givenn in the literature, for example in [5, 6, 9, 13].

Let us consider now a supervisory control problem.In general, the supervisory control goal may be formulated as follows:

Given a set of supervisory control systems controllable by the auxiliary vector u^2, find x^2 and $u^2 = C^2(x^2,t)$ which will minimize the overall performance index I.

If

$$u_o^2 = (\xi_o, \rho_o, \omega_o)$$

where ξ_o, ρ_o, ω_o - optimal interaction vector obtained from the optimal solution of the overall direct control problem, then the two-level control system will be optimal too if and only if a direct control (first level) is optimal.

Since the dimension of each direct optimal controller, due to decomposition, is not large, the complexity of these optimal controllers may be acceptable. However, the complexity of the optimal supervisory controller may be too high, therefore more often we must use sub-optimal solutions for supervisory control. Usually the supervisory controller allocates the resources or interactions to the particular direct control systems, hence the supervisory control problem can be transformed to linear or nonlinear programming problems for which solution algorithms are known [2, 13]. However, since we use sub-optimization, it is usefull to find a few different general forms of sub-optimal control algorithms.

After having found the supervisory control algorithms, we can find the lists of data processing operations which are needed, and define

$$Z(C^2) = Z(n^2), \quad \text{where} \quad n^2 = \dim u$$

Of course, the second-level supervisory control problem may be decomposed into a set of second-level supervisory control problem, and the third-level supervisory control problem has to be introduced.

4. Structure Synthesis

Let us assume that for a given multi-level control problem we have:

a set of direct control algorithms A^1, where $C^1_{i,j} \in A^1$, $j = 0, 1, 2, \ldots$, a complexity function $Z(C^1_{i,j})$ adjoint with with each $C^1_{i,j}$,

a set of second-level supervisory control algorithm A^2, where $C^2_{i,j} \in A^2$, and a set of complexity function $Z(C^2_{i,j})$,

and so on up to a set of k-level supervisory control algorithms A^k, where $C^k_{i,j} \in A^k$ with adjoint complexity functions $Z(C^k_{i,j})$.

These sets of control algorithms can be ordered according to the values of the complexity functions Z for a given dimension of control, such that

$$\langle C^r_{i,0}, Z(C^r_{i,0})\rangle , \langle C^r_{i,1}, Z(C^r_{i,1})\rangle \ldots \langle C^r_{i,j}, Z(C^r_{i,j})\rangle$$

where

$$Z(C^r_{i,0}) > Z(C^r_{i,1}) > Z(C^r_{i,2}) > \ldots > Z(C^r_{i,j})$$

and

$$I(C^r_{i,0}) < I(C^r_{i,1}) < I(C^r_{i,2}) < \ldots < I(C^r_{i,j})$$

where: $C^r_{i,0}$ - optimal control algorithm, $C^r_{i,j}$, $j = 1,2,\ldots,$ - sub-optimal control algorithms.

Each of the complexity functions Z is a function of the dimension of a given control problem, therefore

$$Z_k = Z(C^k_{i,j}, n^k)$$

$$Z_{k-1} = Z(C^{k-1}_{1,j}, n^{k-1}_1) + Z(C^{k-1}_{2,j}, n^{k-1}_2) + \ldots + Z(C^{k-1}_{i,j}, n^{k-1}_i) + \ldots + Z(C^{k-1}_{p,j}, n^{k-1}_p)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$Z_1 = Z(C^1_{1,j}, n^1_1) + Z(C^1_{2,j}, n^1_2) + \ldots + Z(C^1_{i,j}, n^1_i) + \ldots + Z(C^1_{s,j}, n^1_s)$$

where $n^1_{i,j}$ - dimension of the control vector for the control algorithm $C^1_{i,j}$.

Since in the general case we can have several possible control algorithms for each control level, there exists a set of Z function models of the multi-level control system. From this set we must choose the admissible Z function models of the multi-level system and then choose the control structure which can give the most desirable value of the performance index.

The problem of control structure synthesis can be solved by using the following minimization procedures:

By solving the following minimization problem:

$$\underset{\substack{n^2 = \dim u^2 \\ \vdots \\ n^k = \dim u^k}}{\text{minimize}} \; Z_{g\,\min} = \min_{C^k \in A^k} Z_k + \min_{C^{k-1} \in A^{k-1}} Z_{k-1} + \ldots + \min_{C^1 \in A^1} Z_1$$

we can find if an admissible Z function model of the multi--level control structure exists.

An approximate solution of this problem can be easily obtained by assuming that $n^k, n^{k-1}, \ldots, n^k$ are no integers from

$$\operatorname{grad} Z_{g\,min} = \left(\frac{\partial Z_{g\,min}}{\partial n^2}, \frac{\partial Z_{g\,min}}{\partial n^3}, \ldots, \frac{\partial Z_{g\,min}}{\partial n^k} \right)^T = 0$$

and then by choosing the best integer approximation of the solution.

If $\min\limits_{n^i} Z_{g\,min} > Z_o$, the admissible structure does not exist in the given set of multi-level structures, and it is necessary to find and to include in this set new more simple sub--optimal control algorithms, of course if such control algorithms exist.

If $\min\limits_{n^i} Z_{g\,min} = Z_o$, only one admissible multi-level control structure exists, and the structure synthesis problem is solved.

If $\min\limits_{n^i} Z_{g\,min} < Z_o$, in general, several admissible multi--level structures exist, and it is necessary to choose such a structure which gives after parameter optimization the most desirable value of the performance index I .Since the sets of control algorithms are ordered, the solution of the following minimization problem:

$$\underset{\substack{n^2 = \dim u^2 \\ \vdots \\ n^k = \dim u^k}}{\text{minimize}} \; Z_{g\,max} = \max_{C^k \in A^k} Z_k + \max_{C^{k-1} \in A^{k-1}} Z_{k-1} + \ldots + \max_{C^1 \in A^1} Z_1$$

gives the answer, if the multi-level control structure with optimal control algorithms is admissible or not.

If $\min\limits_{n^i} Z_{g\,max} \leqslant Z_o$, the solution of this minimization problem gives us the optimal multi-level control structure, however in general it is not necessaryly the unique optimal control structure.

If $\min\limits_{n^i} Z_{g\,max} > Z_o$ and $\min\limits_{n^i} Z_{g\,min} < Z_o$, there may exist several admissible multi-level structures, however, all these structures guide to some performance index I losses. To minimize these losses we are looking for the most complicated sub--optimal algorithms for which

$$\min_{n^i} Z_g \simeq Z_o \; .$$

<u>Example</u>. Let us consider a very simple example.We first assume that we have found one direct optimal control algorithm C_0^1 with a complexity function $Z(C_0^1) = k_1[(n^1)^2/n^2]$, one optimal supervisory control algorithm for the second level of control C_0^2 with a complexity function $Z(C_0^2) = k_2[(n^2)^3/n^3]$ and one sub-optimal supervisory control algorithm for the second level of control C_1^2 with a complexity function $Z(C_1^2) = k_3[(n^2)^2/n^3]$, one optimal supervisory control algorithm for the third level of control C_0^3 with a complexity function $Z(C_0^3) = k_4(n^3)^3$, one sub-optimal supervisory control algorithm for the third level of control C_1^3 with a complexity function $Z(C_1^3) = k_5(n^3)^2$.

Then we assume that $n^1 = 8$, $k_1 = k_2 = \ldots = k_5 = 1$, $Z_o = 40$, $n^2, n^3 > 1$, and formulate

$$Z_{g\,min} = \min_{C^1 \in A^1} Z_1 + \min_{C^2 \in A^2} Z_2 + \min_{C^3 \in A^3} Z_3 = \frac{8^2}{n^2} + \frac{(n^2)^2}{n^3} + (n^3)^2$$

and find $\min\limits_{n^2,\, n^3} Z_{g\,min} = 28$, where $n^2 = 4$, $n^3 = 2$ (Fig. 2a).

Since $\min\limits_{n^i} Z_{g\,min} < Z_o$, we formulate

$$Z_{g\,max} = \max_{C^1 \in A^1} Z_1 + \max_{C^2 \in A^2} Z_2 + \max_{C^3 \in A^3} Z_3 = \frac{8^2}{n^2} + \frac{(n^2)^3}{n^3} + (n^3)^3$$

and find $\min_{n^2,n^3} Z_{g\ max} = 40$, where $n^2 = 2$, $n^3 = 0$ (Fig. 2b).

Thus we have accepted the multi-level structure of Fig. 2b, however, if $Z_o < 40$, the acceptable multi-level structureis the one presented in Fig. 2a.

References

1. Chang A.: An optimal regulator problem. J. Soc. Ind. Appl. Mathem. Ser. A: Control 1964 No. 2 pp. 220-233.
2. Dantzig C.B.: Linear programming and extension. Princeton 1963.
3. Feldbaum A.A.: Theory of the optimal control systems. Moscow 1963, in Russian.
4. Golubiew-Nowozilow J.S.: Multicomputers complex systems. Moscow 1967, in Russian.
5. Kulikowski R.: Optimum control of aggregated multi-level system. Proc. III IFAC Congr. London 1966.
6. Kulikowski R.: Optimum control od dynamic interacting systems. Arch. Automat. i Telemechan. 1966 No. 2.
7. Lariczew O.I.: Time optimum control of systems connected by limitation. Pap. III IFAC Congr. London 1966.
8. Mesarovic M.D., Lefkowitz I., Pearson J.: Advances in multi-level control. IFAC Tokyo Symp. on System Engineering Tokyo 1965.
9. Mesarovic M.D. et all.: On the synthesis of dynamic multi-level systems. Pap. III IFAC Congr. London 1966.
10. Mesarovic M.D., Macko D., Takahare Y.: Structuring on multi-level systems. Preprints IFAC Symp. on Multi-variable Control Systems Düsseldorf 1968.
11. Pearson J.D.: Multilevel control systems. IFAC Symp. on Adaptive Control Teddington 1965.
12. Straszak A.: On the structure synthesis problem in multi-level control systems. Proc. Tokyo Symp. IFAC on System Engineering for Control System Design Tokyo 1965.
13. Straszak A.: Sub-optimal supervisory control. In: Functional Analysis and Optimization. New York 1966.
14. Straszak A.: Optimal and sub-optimal multivariable control systems with controller cost constraint. Proc. III IFAC Congr. London 1966.

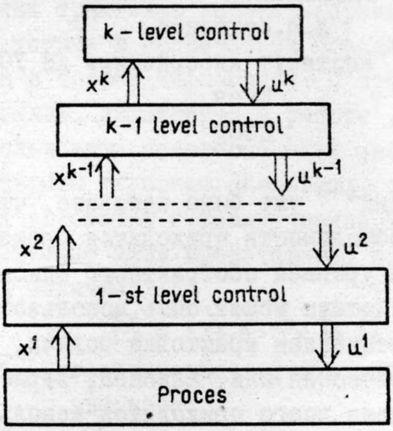

Fig. 1

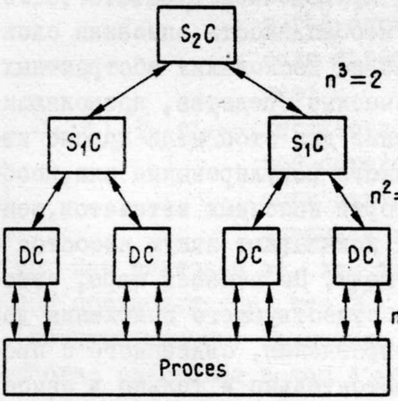

Fig. 2a

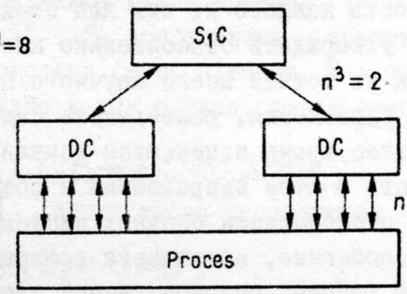

Fig. 2b

О ТЕОРИИ УПРАВЛЕНИЯ СЛОЖНЫМИ СИСТЕМАМИ

А.И.Кухтенко
Институт кибернетики АН УССР
Киев
СССР

В ряде публикаций[I-4] уже было показано, что изучение сложных систем по необходимости приходится производить при использовании различных уровней абстрактного описания. В зависимости от назначения системы может быть использована, например, либо теоретико-информационная трактовка задачи, либо логико-математическая, динамическая или, наконец, эвристическая. В действительности же чаще всего приходится использовать одновременно несколько различных уровней абстрактного описания.

Не останавливаясь здесь повторно на характеристике термина "сложная система управления", приведенной в работе[3], отметим только то обстоятельство, что необходимость описания сложной системы управления одновременно на нескольких абстрактных уровнях заставляет искать математические средства, позволяющие это делать. Однако попытки применения для этой цели хорошо известных методов теории автоматического регулирования или, вообще, теории динамических систем, теории конечных автоматов, шенноновский теории информации и т.д. показывают явную несостоятельность каждого из них для этой цели. По меньшей мере, это можно утверждать относительно ныне существующего положения дел. Каждая из ветвей всего научного направления, связанного с проблемой управления, развивалась самостоятельно и только в самое последнее время намечаются контакты между ними. Описанию известного в этом направлении и обсуждению некоторых возможных путей исследования сложных систем управления и посвящено настоящее сообщение, являющееся реферативным изложением части более полной работы, подготовленной автором для печати по данному вопросу.

§ I. О единой концепции в теории конечных автоматов и динамической теории управления

Вполне естественно, что прежде, чем будет создана "Общая теория управления", которая позволит всесторонне изучать поведение сложной системы управления, различные исследователи пытаются решить более простую задачу, стремятся создать метод, охватывающий одновременно хотя бы только две из числа возможных абстрактных трактовок задач. Например, в работах[5,6] авторы их прилагают усилия в направлении объединения методов динамической теории и теории информации. С этой точки зрения весьма интересны взгляды, изложенные в работе[7], автор которой показал, что если использовать некоторые идеи абстрактной алгебры, то могут быть вскрыты глубокие аналогии, существующие между теорией конечных автоматов и динамической теорией управления. Если конечный автомат Мили представить, как это обычно и делается[8], пятеркой величин:

$$M=\{X, Z, S_\nu, f_z, f_s\},$$

где X и Z – соответственно, входной и выходной алфавиты автомата;

S_ν – множество величин, определяющих состояние автомата;

$Z_\nu = f_z(S_\nu, x_\nu)$ – характеристическая функция, на основании которой определяются выходные величины автомата, если известны входные величины и его состояние;

$S_{\nu+1} = f_s(S_\nu, x_\nu)$ – характеристическая функция, на основании которой определяется состояние автомата в $\nu+1$ такт, если известны входные величины и состояние автомата в ν такт,

то, как показано в[7], для дискретной шкалы времени $T=\{0,1,2,\dots\}$, при предположении, что X, Z, S представимы абелевыми (коммутативными) группами, система (т.е. автомат M) будет вполне аддитивна в том и только в том случае, когда существуют такие гомоморфизмы[9]

$$A: S \to S; \quad B: X \to S; \quad C: S \to Z; \quad \mathfrak{D}: X \to Z;$$

что для всех $S_\nu \in S$, $x \in X$ имеют место соотношения

$$S_{\nu+1} = f_s(S_\nu, x_\nu) = AS_\nu + Bx_\nu,$$

$$Z_\nu = f_z(S_\nu, x_\nu) = CS_\nu + \mathfrak{D}x_\nu,$$

которые можно записать в более развернутом виде

$$S_{\nu+1} = f_s(S_\nu, x_1, \dots, x_n) = A^n S_\nu + \sum_{m=1}^{n} A^{m-1}_{n-m+1},$$

$$Z_\nu = f_z(S_\nu, x_0, \dots, x_n) = CA^{n-1} S_\nu + \sum_{m=1}^{n} CA^{m-1} B x_{n-m} + \mathfrak{D} x_n$$

Если теперь ввести обозначения

$$\Phi(n)=CA^{n-1}$$

$$h(m)=\begin{cases} \mathcal{D} & \text{при } m=0 \\ CA^{m-1}B & \text{при } m>0, \end{cases}$$

то выражение для f_z можно будет записать в следующем виде:

$$z_\nu=f_z(S_\nu,x_\nu,\ldots,x_n)=\Phi(n)S_\nu+\sum_{m=0}^{n-1}h(m)x_{n-m}$$

Заменяя, наконец, дискретное время непрерывным и переходя от гомоморфизма групп к гомоморфизму векторного пространства, последнему выражению можно придать вид, хорошо известный в линейной теории динамических систем управления[10]:

$$z(t)=\Phi(t-t_0)x(t_0)+\int_{t_0}^{t}h(t-\xi)u(\xi)d\xi,$$

где $z(t)$ - P-мерный вектор, характеризующий выход системы;
$x(t_0)$ - h-мерный вектор, характеризующий состояние системы в момент времени t_0;
$u(t)$ - 7-мерный вектор, характеризующий вход системы;
$\Phi(t)-(P\times n)$ - матрица, i-й столбец которой представляет собой реакцию системы в момент времени t;
$h(t)$ - импульсная переходная функция системы.

В возможности такого рода переходов от соотношений теории конечных автоматов к соотношениям линейной динамической теории и проявляется та глубокая связь, которая позволяет говорить о единой концепции для этих двух ранее независимо развивавшихся ветвей научных знаний. Представляется, однако, необходимым сделать еще один шаг вперед по пути объединения логической и динамической трактовок задач теории управления сложными системами, обсуждению возможности чего и посвящен следующий параграф

§ 2. О логико-динамических системах

Не трудно осознать, что кроме демонстрации самого факта существования глубокой связи, обнаруживаемой между теорией конечных автоматов и теорией линейных динамических систем, для фактического изучения сложных систем управления, состоящих одновременно из логических и динамических звеньев, имманентно объединенных в одно целое, требуется построение совершенно новой теории, а для этого требуется также новые математические средства. Основное затруднение, возникающее при изучении логико-динамических систем (как ради краткости будем их именовать), заключается в том, что необходимо найти такой гибкий язык,который позволил бы оперировать одинаково удобным образом как с обычными переменными математического анализа, так и с логическими переменными. При создании такого рода языка, по-видимому, можно идти различными путями. С этой точки зрения заслуживает внимания: язык R-функций[11], операторное исчисление для булевых функций[12], язык непрерывной логики[13] и др. В частности, для этих же целей в работе[14] вводится понятие <u>гибридной функции $G(x_1, \dots, x_e)$</u>, представляющей собой произведение обычной функции действительного переменного - $F(x_i, \dots, x_n)$ и функции логических переменных $f(x_j, \dots, x_m)$, т.е.

$$G(x_1, \dots, x_e) = F(x_i, \dots, x_n) \cdot f(x_j, \dots, x_m)$$

Логическая функция $f(x_j, \dots, x_m)$ может быть предикатом, формулой или квантором, но она может принимать только два значения: I (истина) и 0 (ложь), однако моменты, когда это будет происходить, зависят сложным образом от значения переменных различной природы. Они могут быть:

1) предикатами, зависящими от функций действительного переменного,
2) частично действительными переменными, а частично - логическими переменными,
3) только логическими переменными, не зависящими от действительных переменных.

Все это создает разнообразные возможности при описании логико-динамических систем. К сожалению, сам факт введения гибридной функции еще не означает, что уже имеется необходимый нам язык. Нужны дальнейшие исследования, связанные с разработкой правил, при помощи которых можно было бы производить необходимые действия с гибридными функциями (дифференцировать их, интегрировать и т.д.), поэтому вопрос о создании математического аппарата, пригодного для изучения логико-динамических систем,остается еще открытым. Уже имеются работы, в которых развивается аппарат, основанный на понятии гибридной функции[15], но имеются и критические работы[16], в которых такого рода возможности отрицаются. Сейчас можно только отметить, что известное предсказание Дж.фон-Неймана о необходимости слияния методов непрерывной и конечной математики (т.е. слияния методов математического анализа, опирающихся на факт непрерывности переменных, и методов математической логики, оперирующих с дискретными логическими переменными) сбывается и нужды построения теории логико-динамических систем, по-видимому, будут побуждать многих исследователей скорее идти к этой цели.

Можно также отметить, что не меньшие трудности возникают и при изучении систем, требующих одновременного использования двух любых других абстрактных уровней описания сложных систем, например, логического и эвристического. Еще большие трудности возникают, если используется одновременно не два, а три уровня абстрактного описания систем, например, информационный, логический, эвристический. Создание математических средств, позволяющих вести исследования в таких и еще более сложных случаях, и является главной целью общей теории управления сложными системами.

§ 3. О проблеме многомерности в теории сложных систем

Отмеченные выше трудности не являются единственными, с которыми приходится встречаться при исследовании подобного рода систем. Существенные затруднения возникают и при использовании одного какого-либо уровня абстрактного описания, если система состоит из многих взаимосвязанных друг с другом элементов или подсистем.

При этом "проклятие многомерности" (по образному выражению Р.Беллмана[17]) в одинаковой мере трудно преодолевается при использовании любого из уровней абстрактного описания. Так,например, в теории конечных автоматов задачи легко и изящно решаются при малой размерности систем, и трудности существенно возрастают как только размерность изучаемой системы возрастает. Поэтому вполне естественным путем возникает желание найти такие пути преодоления трудностей, связанных с многомерностью системы, которые, например, были бы одновременно пригодны для изучения многомерных динамических и логических систем большой размерности. Как нам представляется, бороться с многомерностью и многосвязанностью сложных систем можно не только путем изыскания новых математических методов (чем, в основном, только и ограничивались), но и путем использования физических и инженерных сведений о изучаемых системах. Так, можно использовать тот факт, что система (будь она технического, экономического, биологического или социального характера) состоит из больших групп однотипных элементов. Если это так, то становится возможным использовать математический аппарат, разработанный для описания многофазных жидкостей, как это предлагали делать при изучении такого рода систем И.М.Гельфанд и М.Л.Цейтлин[18]. Второй возможный путь преодоления трудностей, связанных с многомерностью задачи, рассмотренный в[4], также базируется на физических представлениях. А именно: когда система в том или ином смысле симметрична, представляются возможности существенно упростить исследование сложной системы большой размерности как при динамической, так и логической трактовке задач. В этих случаях можно воспользоваться методами теории групп[19], а точнее - теории представления групп[20], столь широко применяемыми в квантовой физике, квантовой химии и в современной теории элементарных частиц[21-23].

Не следует при этом полагать, что речь идет непременно о симметрии чисто геометрического характера. Отнюдь нет. Наиболее важными, для интересующей нас проблемы исследования систем большой размерности, являются более глубоко скрытые свойства симметрии. Так, для многих и весьма разнообразных динамических систем свойство симметрии проявляется в том, что для них лагранжиан (или гамильтониан) остается инвариантным относительно линейных преобразований координат. Если же тем или иным способом установлен факт наличия симметрии (например, это может быть непосредственно обнаружено при использовании матричной формы записи уравнений), то формальный аппарат теории представления групп позволяет исходную задачу большой размерности заменить несколькими однотипными задачами значительно меньшей размерности (например, в 10 или в 100 раз меньшей размерности). Суть дела при этом заключается в том, что соответствующая исходная матрица большой размерности может быть по вполне определенной рецептуре, обусловленной видом имеющейся симметрии, сведена к блочно-диагональному виду

$$\begin{Vmatrix} \Gamma_{ij}^{(1)}(P) & 0 & 0 \ldots\ldots & 0 \\ 0 & \Gamma_{ij}^{(2)}(P) & 0 \ldots\ldots & 0 \\ 0 & 0 & \Gamma_{ij}^{(3)}(P) \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 \ldots\ldots & \Gamma_{ij}^{(n)}(P) \end{Vmatrix},$$

где каждая из субматриц, стоящих на главной диагонали,уже является неприводимой, т.е. не может быть еще один раз представлена через матрицы еще меньшей размерности. Теория представления групп позволяет ответить на вопросы о том, какова может быть размерность субматриц $\Gamma_{ij}(P)$, каково их число, сколькими различными способами можно производить такого рода представление матрицы большой размерности с помощью матриц малой размерности, в частности, дает критерии для суждения о том , возможно или нет дальнейшее разложение субматриц $\Gamma_{ij}(P)$ на матрицы еще меньшей размерности и т.д. При кратком изложении нельзя описать все те процедуры, которые приходится при этом практически выполнять: разбить элементы групп на классы, найти характер для каждого класса, т.е. след субматрицы $S_P\|\Gamma_{ij}(P)\|$, определить канонический базис для исходного многомерного векторного пространства данной задачи и т.д. Со всеми необходимыми понятиями и теоремами (эквивалентность и гомоморфизм групп, леммы Шура, теоремы Лагранжа и Вигнера и т.д.) и методикой пользования ими можно ознакомиться по указанным выше книгам , посвященным изложению теории групп и теории представления групп[18-23]. Поскольку исследователю часто представляется возможность самому формировать структуру изучаемой сложной системы (например, при управлении экономическими системами и производственными объектами), то, выбрав структуру симметричной , имеется возможность воспользоваться идеями теории представления групп, в связи с чем последняя приобретает весьма большое практическое значение.

Для пояснения сути дела рассмотрим элементарно простой пример - колебания некоторой механической системы (см. детальнее[4]).

Пусть уравнения, записанные в матричной форме, будут иметь вид:

$$M\ddot{X}+KX=0,$$

где M - диагональная матрица масс,
K - матрица жесткостей,
X - вектор-столбец переменных

$$K = \begin{Vmatrix} \frac{3}{2}K_2 & 0 & 0 & -\frac{\sqrt{3}}{2}K_2 & \frac{\sqrt{3}}{2}K_2 \\ 0 & \frac{3}{2}K_2 & -K_2 & \frac{K_2}{2} & \frac{K_2}{2} \\ 0 & -K_2 & \sqrt{3}K_1+K_2 & 0 & 0 \\ -\frac{\sqrt{3}}{2}K_2 & \frac{K_2}{2} & 0 & \sqrt{3}K_1+K_2 & 0 \\ \frac{\sqrt{3}}{2}K_2 & \frac{K_2}{2} & 0 & 0 & \sqrt{3}K_1+K_2 \end{Vmatrix}, \qquad X = \begin{Vmatrix} x \\ y \\ y_1 \\ y_2 \\ y_3 \end{Vmatrix}$$

$$M = \begin{Vmatrix} m_1 & 0 & 0 & 0 & 0 \\ 0 & m_1 & 0 & 0 & 0 \\ 0 & 0 & m & 0 & 0 \\ 0 & 0 & 0 & m & 0 \\ 0 & 0 & 0 & 0 & m \end{Vmatrix}.$$

Таким образом, необходимо исследовать колебания системы, имеющей матрицы размерностью 5 . Задача заключается в том, чтобы, пользуясь свойством симметрии, вместо исходной задачи можно было бы решить задачу меньшей размерности. Пользуясь известной методикой теории представления групп[18-23] и тем фактом, что исследуемая колебательная система симметрична (что проявляется в симметрии матриц), могут быть найдены новые переменные $h_{11}, h_{12}, h_{21}, h_{22}, h_{23}$, которые связаны со старыми переменными x, y, y_1, y_2, y_3 такими соотношениями, что задача размерности 5x5 распадается на две одинаковые задачи размерности 2x2 и одну задачу размерности IxI. Не излагая здесь всех подробностей преобразований, основанных на идеях теории представления групп, выпишем конечные соотношения, связывающие старые и новые переменные

$$y_1 = \frac{1}{\sqrt{3}}\sum_{\alpha=1}^{\alpha=3} h_{\alpha 1}, \qquad x = \frac{1}{\sqrt{2}}(h_{12}+h_{22}),$$

$$y_2 = \frac{1}{\sqrt{3}}\sum_{\alpha=1}^{\alpha=3} e^{\frac{2\pi i\alpha}{3}} h_{\alpha 1}, \qquad y = -\frac{i}{\sqrt{2}}(h_{12}-h_{22})$$

$$y_3 = \frac{1}{\sqrt{3}}\sum_{\alpha=1}^{\alpha=3} e^{\frac{2\pi i\alpha}{3}} h_{\alpha 1},$$

В матричной форме эти же соотношения могут быть записаны в следующем виде:

$$X = RH,$$

где матрица R и вектор-столбец H имеют вид:

$$R=\left\| \begin{matrix} 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}}e^{\frac{2\pi i}{3}} & 0 & \frac{1}{\sqrt{3}}e^{\frac{2\pi i}{3}} & 0 & \frac{1}{\sqrt{3}}e^{2\pi i} \\ \frac{1}{\sqrt{3}}e^{-\frac{2\pi i}{3}} & 0 & \frac{1}{\sqrt{3}}e^{\frac{2\pi i}{3}} & 0 & \frac{1}{\sqrt{3}}e^{2\pi i} \end{matrix} \right\|, \quad H=\left\| \begin{matrix} h_{11} \\ h_{12} \\ h_{21} \\ h_{22} \\ h_{23} \end{matrix} \right\|$$

Обратная связь новых переменных со старыми в матричном виде записывается в виде уравнения

$$H = R^{-1}X,$$

где R^{-1} - матрица, обратная по отношению к матрице R. Подстановкой линейного преобразования $X = RH$ в исходное уравнение и умножением слева на матрицу R^{-1}, исходная система дифференциальных уравнений приводится к блочно-диагональному виду:

$$R^{-1}MR\ddot{H} + R^{-1}KRH = 0,$$

где $R^{-1}MR$ - диагональная матрица масс вида

$$R^{-1}MR = \left\| \begin{matrix} m & 0 & 0 & 0 & 0 \\ 0 & m_1 & 0 & 0 & 0 \\ 0 & 0 & m & 0 & 0 \\ 0 & 0 & 0 & 0 & m_1 \end{matrix} \right\|,$$

а $R^{-1}KR$ - блочно-диагональная матрица жесткостей вида

$$R^{-1}KR = \left\| \begin{matrix} \sqrt{3}K_1 + K_2 & \frac{\sqrt{3}K_2 i}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & \frac{3}{2}K_2 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{3}K_1 + K_2 & -\frac{\sqrt{3}K_2 i}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{\sqrt{3}K_2 i}{\sqrt{2}} & \frac{3}{2}K_2 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{3}K_1 + K_2 \end{matrix} \right\|$$

Таким образом, на основании вполне определенной процедуры мы пришли к блочно-диагональному виду матриц. В данном методе это делается не на основе метода попыток, а по вполне определенным алгоритмам, если исходная система симметрична. Система дифференциальных уравнений в новых переменных имеет вид:

$$\left.\begin{aligned} m\ddot{h}_{11}+(\sqrt{3}K_1+K_2)h_{11}+i\sqrt{\tfrac{3}{2}}K_2h_{12}=0 \\ m_1\ddot{h}_{12}+\tfrac{3}{2}K_2h_{12}-i\sqrt{\tfrac{3}{2}}K_2h_{11}=0 \end{aligned}\right\}$$

$$\left.\begin{aligned} m\ddot{h}_{21}+(\sqrt{3}K_1+K_2)h_{21}-i\sqrt{\tfrac{3}{2}}K_2h_{22}=0 \\ m_1\ddot{h}_{22}+\tfrac{3}{2}K_2h_{22}+i\sqrt{\tfrac{3}{2}}K_2h_{21}=0 \end{aligned}\right\}$$

$$m\ddot{h}_{31}+(\sqrt{3}K_1+K_2)h_{31}=0$$

Таким образом, наглядно видно, что вместо решения задачи размерности 5x5, дело сводится к двухкратному решению одной и той же задачи размерности 2x2 и отдельному решению уравнения размерности IxI.

Мы интересуемся проблемой управления сложными системами, а изложенный только что метод дает возможность исследовать только отдельно взятый объект управления, имеющий симметричную структуру, если система управления не симметрична. Поэтому в такого рода случаях задачи устойчивости, инвариантности, оптимальности и др. являются самостоятельными задачами, каждая из которых решается на основе использования свойств симметрии объекта. Так, задача устойчивости может быть решена путем использования метода декомпозиции, с помощью которого вся система расчленяется на подсистемы. В первую из этих подсистем входит объект, описываемый симметричной матрицей, а во вторую подсистему - система управления. Первоначально исследуется устойчивость каждой из подсистем, причем при изучении устойчивости объекта, взятого отдельно, используются свойства симметричности его. Определив устойчивость подсистем, затем, на основании, например, известной теоремы Бейли, могут быть определены условия асимптотической устойчивости для всей системы в целом. Особо интересными являются задачи оптимального управления для такого же рода объектов, когда совместное использование метода декомпозиции и идей теории представлений групп, позволяет решать весьма трудные многомерные задачи оптимального управления. Фактическая демонстрация подобного рода возможностей требует отдельного сообщения.

§ 4. Об универсальности ЭЦВМ и моделировании процессов управления сложными системами

Главная особенность теории сложных систем, как и кибернетики вообще, проявляется в том, что она должна давать возможность изучать объекты любой природы (технические, экономические, биологические, социальные) на соответствующем абстрактном уровне. Поэтому необходимо заботиться не столько о создании "Общей теории систем", из которой как частные случаи получались бы частные теории (теория линейных динамических систем, теория информации, теория марковских процессов и т.д.), к чему иногда призывают[24], сколько о создании многих ветвей теории для различных уровней абстрактного описания систем. Рассмотрение задачи на каком-либо одном уровне абстракции позволяет дать ответы только на определенную группу вопросов, а для получения ответов на другие вопросы необходимо произвести исследования на другом уровне абстрактного описания системы. Для достижения максимально возможной полноты необходимо изучение одной и той же системы на всех подходящих для данного случая уровнях абстракции. Наиболее целесообразным для этой цели и практически доступным является путь математического моделирования. Хотя обычно на ЭЦВМ смотрят просто как на весьма быстродействующий вычислитель, в действительности же она является универсальным устройством, которое кроме быстрого счета способно производить обработку буквенных или других символов, преобразовывать информацию в необходимую нам форму, делать "умозаключения", "выводы" и т.п. Все это позволяет с помощью ЭЦВМ изучать не только информационные процессы в сложных системах управления, но и логические, динамические и эвристические трактовки задач также могут быть обследованы. В настоящее время созданы специальные абстрактные языки (Симскрипт, SIMPAC, GPSS и др.[25]), позволяющие экономить время и усилия, связанные с программированием и самим процессом моделирования. Поэтому именно этот путь использования ЭЦВМ для моделирования и является в настоящее время главным при разработке действительно сложных систем управления. Часто также идут на создание специальных научных центров, предназначенных исключительно для целей моделирования разрабатываемой сложной системы управления. В качестве примера можно привести научный центр, созданный специально для разработки системы автоматического управления воздушным движением над территорией Западной Европы[26]. Чтобы охарактеризовать действительную сложность подобного рода систем управления, можно привести данные о программах, необходимых для управления с помощью ЭЦВМ и других технических средств (радиолокационных станций, связного оборудования и т.д.) потоками самолетов (общим числом 300-600) над территорией , протяженностью около 2000 км[27]. Только для разработки таких программ необходимы усилия 250 программистов в течение 2-х лет и общее количество команд в этих программах составляет 2,5 млн. Не смотря на то, что стоимость такого рода научно-исслловательских моделирующих центров достаточно высока, экономическая целесообразность их создания, при разработке действительно сложных систем управления, несомненная и по этому пути идут во многих случаях как при решении технических, экономических или оборонных задач, так и при выполнении крупных социологических исследований. Поэтому столь значимой становится проблема разработки общей теории преобразования алгоритмов (акад.В.М.Глушков[28]) и вообще теории управления сложными системами, об отдельных аспектах которой выше шла ресъ, так как только при наличии ее в полной мере могут быть использованы столь мощные средства математического моделирования. В настоящем обзоре отмечены, конечно, только некоторые из общих вопросов теории сложных систем и не нашли отражения очень многие важные аспекты ее, связанные с проблематикой надежности сложных систем[29-30], эффективности их[31,32] и т.д.

ЛИТЕРАТУРА

1. Общая теория систем, изд-во "Мир",М.,1966.
2. Акад.А.И.Берг,Б.И.Черняк, Информация и управление, изд-во "Экономика", М.,1966.
3. А.И.Кухтенко, Основные направления развития теории управления сложными системами, в сб."Сложные системы управления", вып.4, изд-во "Наукова думка", Киев,1968.
4. А.И.Кухтенко, Основные задачи теории управления сложными системами, тр.семинара "Сложные системы управления", вып.I, изд-во ИК АН УССР,Киев,1968.

5. А.А.Красовский, Изменение энтропии непрерывных динамических систем, "Известия АН СССР", "Техническая кибернетика", №5, I964.
6. Р.Л.Стратонович, Экстремальные задачи теории информации и динамическое программирование, "Известия АН СССР", "Техническая кибернетика", №5,I967.
7. M. Arbib. A common framework for automata theory and control theory. J. Soc. Industr. and Appl. Math. 1965, A.3 №2.
8. А.Гилл, Введение в теорию конечных автоматов, изд-во "Наука", I966.
9. Р.Фор,А.Кофман,М,Дени-Папен, Современная математика", изд-во "Мир", I966.
IO. В.М.Броун, Анализ линейных инвариантных во времени систем, изд-во "Машиностроение", I966.
II. В.Л.Рвачев, Геометрические приложения алгебры логики, изд-во "Техника", Киев, I967.
I2. I. Richalet. Calcul opérationnel Booleen. "L'onde Electrique", 1965, №439, v43., p63.
I3. С.А.Гинзбург, Непрерывная логика и ее применение, АИТ, №2,I967.
I4. Р.Терно, Гибридные функции - новый метод описания сложных систем, "Известия АН СССР", "Техническая кибернетика", № 6, I965.
I5. М.М.Гердов, Один метод численного анализа сложных систем, "Известия АН СССР", "Техническая кибернетика", № 6,I967.
I6. Г.П.Сбродов, Логические отношения и описания сложных систем управления, тр.конф.по экономической кибернетике, Батуми, I966.
I7. Р.Беллман, Направление математических исследований в теории нелинейных цепей, сб."Математика",8,I5, изд-во ИЛ, I964.
I8. И.М.Гельфанд, М.Л.Цетлин, Онекоторых способах управления сложными системами, "Успехи математических наук", т.I6, вып.I,I962.
I9. А.Г.Курош, Теория групп, изд-во" Наука", I967.
2O. Ф.Д.Мурнаган, Теория представления групп, ИЛ, I95O.
2I. Е.Вигнер, Теория групп и ее применение в квантовой механике", ИЛ,I947.

22. В.Хейне, Теория групп в квантовой механике, ИЛ, 1963.
23. Г.Я.Любарский, Теория групп и ее применение в физике, изд-во Физматлит, 1958.
24. М.Месарович, Основания общей теории систем, сб."Общая теория систем", изд-во "Мир",1966.
25. Г.Марковиц и др. Симскрипт, Алгоритмический язык для моделирования, изд-во "Советское радио", 1966.
26. Control. 11, № 105, 1967.
27. T.B. Steel. The development of vary large programs, "Proc. IEI P. Congr." № 1, 1965, Vol. Spartan Books Washington, 1965.
28. В.М.Глушков, Перспективы автоматизации проектирования вычислительных машин, "Вестник АН СССР",№4,изд-во "Наука",1967.
29. И.П.Попчев, Структура на големите системи - основни зависимости и методика за записване, "Техн.мисъл",№1,1967
30. И.Попчев,М.Панова, Определяне на показателя за эфективност на големи системи, "Извести на института по техническа кибернетика",т.VI,1967.
31. Б.С.Флейшман, О живучести сложных систем, "Известия АН СССР", "Техническая кибернетика", №5,1966.
32. Б.С.Флейшман, Статистические пределы эффективности сложных систем, сб."Прикладные задачи технической кибернетики", изд-во "Советское радио", 1966.