053

The Optimal Regulation of Large-Scale Interconnected Linear Systems

by

John Mathews

A Thesis submitted for the Degree of Doctor of Philosophy

April 1980

Department of Computing and Control
Imperial College of Science and Technology
University of London
180 Queens Gate
London SW7 2BZ

Abstract

This thesis is concerned with the optimal regulation of large-scale interconnected linear systems using multilevel coordination schemes that allow parallel computation. Existing schemes are reviewed, and generalized, and new schemesintroduced, all within the context of the LQ regulator, both in its steady-state and dynamic formulations. This is analogous to the well-known procedure in mathematical programming of discussing the performance of different algorithms with respect to the canonical quadratic programming problem.

It is shown that decomposition schemes arise principally from (1) manipulation of the (Euler)-Lagrange necessary conditions of optimality, where interconnection constraints are adjoined to the objective cost function using multipliers and the entire set of equations is solved in multilevel fashion; (2) from the application of gradient-based schemes to the objective function when system and interconnection constraints have been substituted.

Specific update formulae for all schemes are derived, and the fundamental control sequences generated by different methods are compared. In particular , the well-known Takahara scheme is generalized as a successive approximations procedure, and its convergence over a bounded time interval is established.

Acknowledgements

I should like to thank my supervisor, Dr. J.C. Allwright, for his unfailing help and advice throughout the writing of this Thesis.

I should also like to thank Professor D.Q. Mayne for his encouragement and financial assistance in my attending the Third Kingston Conference on Differential Games and Control Theory, Rhode Island, in June 1978.

Financial support throughout the PhD has been provided by the Science Research Council.

I am grateful to Mrs. D. Abeysekera for typing the manuscript.

To Anna and Dominic

Contents

			Page
List of symbols used			1
Chapter 1		Introduction	
	1.1	General background	4
	1.2	Scope of the thesis	7
	1.3	New contributions of the thesis	10
Chapter 2		General Results and Existing Procedures in the Theory of Decentralization	
	2.1	Introduction and outline of chapter	15
	2.2	Interconnected system model	16
	2.3	Coordination principles	19
	2.4	Interaction operators	27
	2.5	Primal and dual techniques	35
	2.6	Unified approach to decomposition schemes	43
Chapter 3		The steady state interconnected linear quadratic regulator	
	3.1	Introduction and outline of chapter	48
	3.2	Formulation of model and derivation of optimality conditions	50
	3.3	Multilevel schemes derived from Lagrangian theory	55
	3.4	Supremal coordination schemes : decomposition based on successive approximations	68
	3.5	Gradient methods and interaction prediction	77
	3.6	Relations between gradient and Lagrangian schemes	93
Chapter 4		The dynamic interconnected linear quadratic regulator:differential description	
	4.1	Introduction and outline of chapter	98
	4.2	Formulation of model and derivation of Euler-Lagrange equations	99
	4.3	Multilevel solutions of Euler-Lagrange equations	103

Contents

			Page
Chapter	4.4	Successive approximations iteration on Euler-Lagrange equations	108
	4.5	Successive approximations to differential systems	112
	4.6	Proof of convergence of successive approximations scheme	118
	4.7	Applicability and numerical computation of successive approximations scheme	123
Chapter	5	The dynamic interconnected linear quadratic regulator : operator formulation	
	5.1	Introduction and outline of chapter	126
	5.2	Operator formalism in Hilbert space setting	126
	5.3	Gradient-based algorithms	132
	5.4	Interaction prediction scheme	140
	5.5	Successive approximations and multilevel schemes	149
Chapter 6		Concluding remarks	
Bibliography			154

List of symbols

u, U	control, set of inputs
x, X	state, set of states
$P : U \to X$	system input-output map
v, V	performance value, set of values
$G : U \times X \to V$	performance function
$J : U \to V$	cost function
u_i, U_i	local control, set of local controls
x_i, X_i	local state, set of local states
z_i, Z_i	interaction, set of interactions
$P_i : U_i \times Z_i \to X_i$	local system input-output map
$H_i : U \times X \to Z_i$	local coupling map
$H : U \times X \to Z$	aggregate coupling map
$K : U \to Z$	direct aggregate coupling map
$\bar{P} : U \times Z \to X$	decoupled system map
$\bar{G}_i : U_i \times X_i \times Z_i \to V_i$	local performance function
$\bar{J}_i : U_i \times Z_i \to V_i$	local cost function
$\bar{G} : U \times X \times Z \to V$	aggregate local performance function
$\bar{J} \quad U \times Z \to V$	aggregate cost function
α	model coordination input
β	goal coordination parameter
e	coordination input error
$\hat{u}$	local optimal control
u^*	global optimal control
$\psi_i^o : U_i \to V_i$	modified local cost function
$\psi^o : U \to V$	aggregate modified local criterion
$\gamma : U \to V$	goal interaction operator

List of symbols (cont)

$\Gamma : U \to V$	linearized goal interaction operator
$\Gamma_i : U_i \to V_i$	local linearized goal interaction operator
λ, X^*	dual variable, dual state space
μ, Z^*	dual variable, dual interaction space
$L : X \times U \times Z \times Z^* \to V$	Lagrangean with respect to interconnection constraint
$\bar{L} : X \times U \times Z \times Z^* \times X^* \to V$	decentralized Lagrangean with respect to system and interconnection constraints
v	coordinating variable
w	local variable
$V_* = \{v \in X: P(\hat{u}(v)) \in X\}$	
$V_o = \{v \in X: U_i(v) \neq \emptyset \; \forall i = 1,\ldots,N\}$	
$K : U \to \mathbb{R}$	Cohen's core functional (used only in Chapter 2, §6; cannot be confused with direct aggregate coupling map)
$G^v : U \to \mathbb{R}$	Cohen's auxiliary functional
S	large scale system
S_i	ith subsystem
n	dimension of state space
m	dimension of control space
q	dimension of interconnection space
(A,B)	aggregate regulator
(A_i,B_i)	regulator subsystems
(G,H)	system interconnection matrices
(Q,R,S,b)	aggregate system cost
(K,E)	successive approximations matrices (operators)
g(u)	system gradient
s	current direction in conjugate gradient algorithm
$\hat{g}(u,z)$	'decomposed' system gradient
$\hat{\lambda}$	'decomposed' system costate variable

List of symbols (cont)

$H(x,u,\lambda,t)$	system Hamiltonian
$\Phi(t,\tau)$	system transition matrix
$f(t)$, $\phi(t)$, $K(t,\tau)$, λ	Mikhlin and Smolitskiy notation for initial value problem
P,Q,R,h	operators used in proof of differential successive approximations scheme
z,C,D	vectors and matrices used in proof of differential successive approximations scheme
$P(t)$	matrix solution of Riccati equation
$\Phi(t)$	matrix solution of 'tracking'equation
$W(t,\tau)$	system convolution kernel
$\hat{W}(t,\tau)$	decoupled system convolution kernel
$\hat{\Phi}(t,\tau)$	decoupled system transition matrix
$Q,R,W,\hat{W},\Phi\ \hat{\Phi}$	operators used in Hilbert space representation of LQ regulator; not to be confused with corresponding matrix quantities used in static and differential representations.
$h(t)$	solution to Allwright 'simple model' scheme
$p(t)$	costate function used in Allwright 'simple model' scheme

Chapter 1

General Background and Introduction

§1. General background

§2. Scope of the thesis

§3. New contributions of the thesis

1. General background

For most of recorded history, humankind's scientific activity has been directed at the control and regulation of the natural world. The richness and variety of this task has led to the development of a number of scientific disciplines, with the language of mathematics as their common core. But with the Industrial Revolution, and the rise of vast technical systems tapping new sources of energy and operating in unfamiliar environments, engineers have been faced with a new challenge : that of regulating a world of artifacts which is every bit as complex as its natural counterpart.

Control theory is one of the disciplines being drawn upon in this most urgent of tasks. Dating from the time of Maxwell and his investigations into the operation of Watt's governor on steam engines (1852), the theory flowered in the 1940's with the possibility of realizing electrical servomechanisms, stabilizers and other common control configurations. With linear transform theory as its theoretical basis, this 'single-input single-output' control theory generated a wide range of design procedures and tools. It was the advent of the computer however that first created the possibility of applying it to technological systems. These are, at the least, multi-input, multi-ouput entities like jet aircraft, high speed trains or nuclear power stations; more challengingly they are complex systems with a number of interdependent parts, like air traffic control systems.

These new concerns have entrained a remarkable series of theoretical advances, making control theory one of the most intellectually exciting fields of the 60's and 70's. Firstly there has been the recasting of the fundamentals of system theory, which was carried through in the 60's and is associated with the work of such outstanding figures as Kalman, Bellman and Pontryagin, and has continued on a number of

different fronts in the 70's.

Secondly there has emerged a novel concern with the complications of large-scale systems, which have a number of levels of control and a number of control stations, and a number of conflicting optimality criteria. In this sense they extend well beyond the purview of classical single-input single-output linear single-station control. This novel concern has been manifested in the development of new control strategies, based on decomposition, decentralization and hierarchical control.

Decentralized control has emerged as a technique for a number of reasons. Among them are the following

(1) Classical single-station controllers have been applied only to systems of relatively small size. Computational procedures become prohibitively expensive as the number of dimensions rises. Distributing the burden of computation and reducing the need for information flow, promises to render this problem tractable.

(2) A number of practical large-scale systems occur naturally in interconnected form, so that decentralized control of each subsystem, with some form of coordinator taking account of interactions, becomes feasible. Chemical plants, for instance, consist of a number of interconnected reactors and vessels with the inter connections being physically realized through pipes.

(3) The emergence of microprocessors offers a powerful stimulus to the development of distributed processing. This much-used term means different things to different people, and to commercial

computer suppliers it raises problems of software compatibility and the design of interface protocols to allow units to communicate with each other. But more fundamentally, such arrays of microprocessors demand the formulation of new, essentially cooperative, or parallel, computational procedures, and this too is a challenge to be met by modern control theory.

As yet there is no single theoretical framework in which all aspects of complex control can be discussed. Many schools of thought are emerging : in the Soviet Union much effort has been expended on developing a comprehensive theory of cybernetics out of automata theory, and in particular from systems of stochastic automata. There is also the highly idio syncratic approach associated with Malishevski [29] and Opoitsev [28], who develop a general schematisation of the dynamics of collective systems. One promising line of advance, associated with the work of Ho and others, is the development of a multilevel, multiagent, multicriteria optimization theory. Such an enterprise has been subjected to sharp criticism, e.g. Mayne [52] , but optimality will perforce remain a salient characteristic of complex systems along with newer concerns with stability, modularity and 'comprehensibility'. This emerging theory is a distillation of differential games theory, which captures the multi-agent aspect and the involvement of multiple criteria; standard optimization theory; and new multilevel, or hierarchical methods for solving the standard optimization problems. This latter school of thought has generated a number of practical procedures for regulating systems of very large dimension, for which the standard centralized schemes would be quite infeasible.

The first efforts in this direction were the matrix inversion procedures of Kron [1] in 1953, and the Dantzig-Wolfe [2] decomposition schemes for Linear Programming problems in 1960. A general formulation of decomposition and coordination strategies was developed by Mesarovic, Macko, Takahara, Lasdon and Pearson [3,4,5] at the Systems Research Center, Case Institute of Technology, which extended the work on static systems to incorporate dynamic optimization problems. This general formulation has not been too successful in generating original regulation schemes, although it has been taken up in France and used effectively by Cohen, Benveniste, and Bernhard [6,7] at the School of Mines in Paris. Other current centres of work in dynamic decomposition and coordination techniques are LAAS at Toulouse (Singh, Titli et al) [8], the Systems Theory Laboratory at Helsinki (Hakkala, Hirvonen) [9], and the Control Engineering Group at Cambridge University (Javdan, Richards, Simmons, White et al) [10,11], while much work has been done in developing steady-state algorithms by Polish groups under the influence of Professor Findeisen [12] .

§2 Scope of the thesis

Despite the large number of papers published on decomposition schemes for optimal control problems, a sample of which are listed in the Bibliography, very little is still known about the underlying connections and comparative performance of these schemes. Much attention has been devoted to the *derivation* of schemes - but very little to their analysis. Indeed many schemes are proposed [33,45] without any regard paid to their applicability, their convergence, or even their numerical characteristics - suffice that they generate a decomposed

structure of regulation. Fortunately such proposals are in a minority and there are many outstanding papers which do examine all these aspects. Nevertheless there is scope for a look at the underlying characteristics of the schemes proposed so far. It turns out that many of them arise from tackling the necessary conditions of optimality in various well-known ways, such as by use of modified Newton-Raphson techniques, or simple successive approximations. Recognition of this fact both clarifies the nature of the schemes themselves - they previously existed in a kind of theoretical limbo , save for schemes based on solution of the dual - and makes analysis of their applicability and convergence much easier. Indeed the best known and oldest of the decomposition schemes,Takahara's procedure for the LQ problem, is still without a general convergence proof.

Decomposition schemes arise principally from manipulation of the Euler-Lagrange necessary conditions of optimality,where interconnection constraints are adjoined to the objective cost function using multipliers, and the entire set of equations is solved in a multilevel fashion; or from the application of gradient-based schemes to the objective function when system and interconnection constraints have been substituted. This latter is a relatively new area for decomposition schemes, and is systematically explored here for the first time.

In this work the LQ problem is taken as canonical in the same way that the quadratic programming problem is taken as canonical for the purpose of investigating many Mathematical Programming algorithms. Specific formulae are derived for the various published schemes in the context of the single LQ model (both in its steady state and dynamic formulations) and thus systematically compared for the first time. The next step along

this road would be a thorough study of convergence rates; this work remains to be done.

The Thesis is organized into 4 chapters, each one based on a different mathematical representation of the interconnected system. The first model is the general formulation of multilevel systems developed by Mesarovic and co-workers. Using this model the general principles and procedures of decentralization and decomposition are expounded. The second model is a formulation of the static, or steady-state, linear-quadratic regulator. A range of multilevel schemes is introduced in this context, as well as a successive approximations procedure that generalizes the Takahara approach; a number of gradient-based procedures; and the interaction prediction approach with modification of subsystem goals by interaction operators. Because of the analytical simplicity of the static formulation, a number of connections between these different approaches are established directly. The third model is the dynamic interconnected LQ regulator in its conventional differential description. In this context a number of multilevel schemes are reviewed, and the Takahara scheme is generalized as a successive approximations iteration of the Euler-Lagrange equations. Convergence is proven over a bounded time interval of control, thus establishing the validity of the successive approximations approach. The fourth model is the same dynamic interconnected LQ regulator in a Hilbert space setting, described with the aid of operators. This reduces the system description to algebraic formulations once more, as in the static case, and the same procedures - multilevel schemes, gradient schemes, and interaction prediction using interaction operators - are developed. An operator version of the Takahara scheme is derived.

§3. New Contributions of the thesis

(i) Chapter 2 is devoted to a review of existing procedures in the theory of decentralization and decomposition. The model expounded is derived from Mesarovic et al [20] but a novel notation is adopted which accords better with later applications of the model to control systems. In §3 the goal interaction operator is expounded in an original way, and an example due to Cohen and Bernhard [7] is greatly extended. In §5 primal and dual techniques are discussed, and conflicting views cited from the literature. The general approach adopted, namely examination of the necessary conditions of optimality derived from an "extended" Lagrangian, while not original, is exploited to give novel insights. The work of Findeisen et al on primal schemes is reviewed, translated into present notation, and a novel comparison with Mesarovic's notions of applicability and coordinability is offered. Finally in §6 Cohen's unified approach to decomposition schemes is reviewed, but nothing original added.

(ii) Although the concept of the steady-state linear quadratic regulator is certainly not novel, the precise representation of the regulator given in Chapter 3 §1 has not previously been published in the literature. Hence all derivations based on this model are the work of the author. Firstly the necessary conditions of optimality are derived, using Lagrangian techniques. A range of 2-level schemes for solving these conditions is then reviewed; some of these lead to useful decom-

position schemes, but some, unexpectedly, lead to indeterminate systems of equations. Such anomalies have not been mentioned before in general treatments of multi-level schemes.

In §3 possible supremal coodination strategies are discussed. "Reinjection" schemes lead to an original static formulation of the "Takahara" algorithm. This scheme is then generalized, and it is shown how it may be derived from the application of Jacobi iteration to the set of necessary conditions. Specifically if these conditions are described by the linear system

$$Cv = b \tag{1}$$

then we 'approximate' the matrix C by an easily invertible matrix K with error E

$$C = K+E \tag{2}$$

and hence derive the successive approximations scheme

$$v^{k+1} = -K^{-1}E\, v^{k} + K^{-1}b. \tag{3}$$

It is shown that a particular choice of K gives the Takahara scheme . On the other hand, different choices of K give schemes which have never been discussed before in the literature: we mention the cases $K = I$, $K = \gamma I$, and an interesting diagonal procedure, $K = \begin{pmatrix} A & 0 \\ 0 & A^T \end{pmatrix}$ where A is the local system matrix. This leads to a wholly decentralized optimization scheme not dependent on any form of coordination, Convergence conditions for all these schemes are discussed.

It is found that the convergence condition for the Takahara scheme admits of no simple interpretation in terms of system

parameters, but that the wholly decentralized scheme converges if amongst other things a simple "diagonal dominance" condition is met.

The question of accelerating convergence via the Gauss-Seidel iteration scheme is then discussed, apparently for the first time in the multilevel systems literature. The accelerated iteration formulae for the wholly decentralized scheme are derived; then it is shown that the Takahara scheme is already in accelerated format. This striking observation forms another, previously un-noted, attractive feature of the scheme.

In §4 a comprehensive discussion of the application of gradient-based algorithms is given, with the aim of deriving decomposition schemes. Although Cohen [7] has shown that his algorithm is an instance of the modified Newton-Raphson algorithm, such a general gradient-based approach has not been previously attempted in the literature on decomposition.

A general Newton-Raphson $NR(K^{-1})$ scheme is discussed (identical to the contraction map $CM(K^{-1})$ scheme in the case of quadratic costs), and a decomposition scheme based on the conjugate-gradient (K^{-1}) algorithm derived. This builds on Allwright's notion of using simple models to solve complex systems.

In §5 the application of interaction operator is discussed. An algorithm given by Cohen [7] for the dynamic regulator is rederived in the context of the static model. It is then shown *directly* that the fundamental control sequences generated by this algorithm, and the $NR(K^{-1})/CM(K^{-1})$ algorithm discussed in §4, are identical. A novel interpretation of the Takahara

scheme as a decentralized approximation of this algorithm is then discussed.

(iii) In Chapter 4 the standard LQ regulator in its differential description is given, but with the added complexity of a(linear) coupling function, or interconnection constraint. The model is formulated in sufficient generality to encompass all previous models of this system published in the literature. Hence all derivations from this model are again the work of the author.

In §2 the Euler-Lagrange optimality conditions are derived, and in §3, are solved using 2-level procedures. In spite of the decade of intense discussions of decomposition procedures the update formulae for these schemes have not previously been published. Again it is found that several schemes suggested by general approaches fail in this concrete (and most basic) instance. In §4 coordination strategies are discussed, and the Takahara "reinjection" scheme is derived as a specific instance of a general successive approximations solution of the 2-point boundary value problem that constitutes the E-L necessary conditions of optimality. Whereas in Chapter 3 the successive approximations procedure was a straightforward application of Jacobi iteration, the procedure in this chapter is anything but straightforward, and indeed never seems to have been referred to before in the literature. Firstly the application of successive approximations to initial value problems and integral equations is reviewed in §5 , and a scalar example due to the author computed to test speed of convergence. Then the generalization of the Takahara scheme, namely the iterative scheme

$$\dot{v}^{k+1} = Kv^{k+1} + Ev^{k}, \quad v_1(t_o) = v_{1o}, \quad v_2(t_f) = 0 \tag{4}$$

applied to the 2PBVP with split boundary conditions :

$$\dot{v} = Cv, \quad v_1(t_o) = v_{1o}, \quad v_2(t_f) = 0 \tag{5}$$

is then shown to converge to the required solution.

(iv) In Chapter 5 the interconnected LQ regulator is recast from its standard differential description into the language of operators on a Hilbert control space. Although the idea of casting control problems in operator form is certainly not new, the systematic presentation of the LQ regulator in operator format is not yet available in the literature, so this model and its derivations is again largely the work of the author (with considerable help from lecture notes by J.C. Allwright). Cost and gradient formulae are derived in §2, then in §3 a range of gradient-based optimization algorithms is discussed with a view to their generating decomposition schemes. Again, as in the case of the static regulator, such an approach to decomposition has not previously been published. Steepest descent is ruled out, but a successful scheme is derived from $NR(K^{-1})/CM(K^{-1})$ algorithm, as well as from the $CG(K^{-1})$ algorithm. In both these cases the notion of 'simple model' is used.

In §4 an interaction prediction algorithm using interaction operator modification is derived, based largely on the work of Cohen [6]. It is then shown *directly*, by looking at the fundamental control sequences generated by each algorithm, that

this scheme is identical to the NR(K^{-1}) scheme derived in §3. Finally in §5 an operator version of the Takahara scheme is derived, and again generalized as Jacobi iteration applied to the operator system constituting the necessary conditions of optimality.

CHAPTER 2

Generalization Results and Existing Procedures in the Theory of Decentralization

§1. Introduction

§2. Interconnected system model

§3. Coordination principles

§4. Interaction operators

§5. Primal and dual techniques

§6. Unified approach to decomposition schemes

§1. Introduction

This chapter is devoted to a review of the current state of the art in the theory of decentralization and decomposition. The theory is expounded using the general multilevel systems model developed by Mesarovic and co-workers. In §2 this model is formulated in notation that conforms well with standard optimal control notation. In §3 two basic coordination principles are introduced - the interaction prediction principle and the interaction balance principle. Conditions under which these principles may be expected to work are then discussed, and an extended example is worked through. In §4 interaction operators are introduced, and it is shown in an original way how they may be used to adapt a problem to make a coordination principle valid. An example due to Cohen and Bernhard (7) is worked through and then greatly extended. In §5 primal and dual techniques are discussed, and fitted into the general framwork introduced in earlier sections. The approach adopted is to examine the necessary conditions of optimality derived from an 'augumented' Lagrangian; this approach is not original, but it is exploited here in a way to give novel insights. The work of Findeisen et al on primal schemes is reviewed, translated into present notations, and a novel comparison with Mesarovic's notions of applicability and coordinability is offered. Finally in §6 Cohen's unified approach to decomposition schemes is reviewed, again translated into present notation.

§2. Interconnected system model

In this first section we develop the formalism introduced by Mesarovic and his group at the Systems Research Center, Case Institute of Technology, adapting it to our own purposes [20]. In particular we shall develop the important concepts of coordinability [38] and applicability [39] which arise in any discussion of co - ordination schemes for multilevel systems :

Following [39] we depict the multilevel control system as in fig 1, which is a 2-level system with N first level controllers :

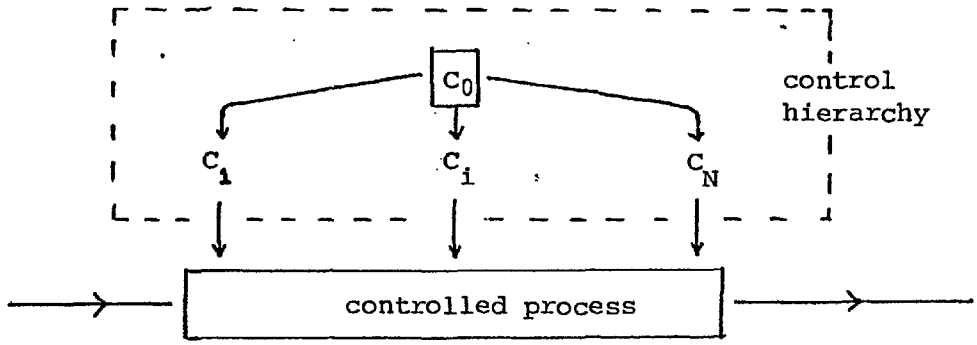


fig 1

We consider a general *process* designated by a map $P : U \to X$, where U is the set of all inputs, and X is set of states (or outputs), so that

$$x = P(u), \quad x \in X, \; u \in U. \tag{1}$$

Attached to this process is a *performance* function $G : U \times X \to V$, where V is the set of performance values,

$$v = G(u,x), \; x \in X, \; u \in U. \tag{2}$$

We consider too the *cost function* $J : u \longmapsto J(u)$ defined by the equation

$$J(u) = G\,[u, P(u)] \tag{3}$$

The overall control problem $\mathcal{P}$ is to find a control u in U which minimises J over U.

We now consider the decomposed system. Physical decomposition will be modelled by mathematical decomposition of the vector spaces U, X into Cartesian products :

$$U = U_1 \times \ldots \times U_N$$

$$X = X_1 \times \ldots \times X_N,$$

so that $u = (u_1, \ldots u_N)$, $u \in U$, $u_i \in U_i$, and similarly for x.

We can now consider the subsystems as *local processes* P_i:

$$U_i \times Z_i \to X_i \quad , \text{ so that}$$

$$x_i = P_i(u_i, z_i) \qquad i = 1, \ldots N, \tag{4}$$

where

$$Z = Z_1 \times \ldots \times Z_N$$

is a newly introduced set of interaction terms.

To specify the *form* of the interaction between subsystems we need a *coupling function*, or *interaction constraint* $H_i : U \times X \to Z_i$, so that

$$z_i = H_i(u,x), \quad i = 1 \ldots N \tag{5}$$

and overall

$$z = H(u,x). \tag{6}$$

Repressing the dependence of z on x, we have

$$z = H(u,P(u)) = K(u). \tag{7}$$

Utilizing the interaction set Z we can now view the system P as being decomposed into a family of *decoupled processes* $\bar{P} : U \times Z \to X$, so that

$$\bar{P}(u,z) = (P_1(u_1,z_1), \ldots P_N(u_N,z_N)). \tag{8}$$

The process $\bar{P}$ is an entirely general construction, and its outputs will coincide with that of the coupled process P if the

the interaction constraint (5) is satisfied. We express this in terms of the equivalence

$$P(u) = \bar{P}(u,K(u)). \tag{9}$$

Hence we have three maps with which to discuss decomposition : the overall process P; the decoupled process $\bar{P} = (P_1,\ldots P_N)$; and the interaction, or coupling or coordination map K(or H).

We depict this situation in fig 2.

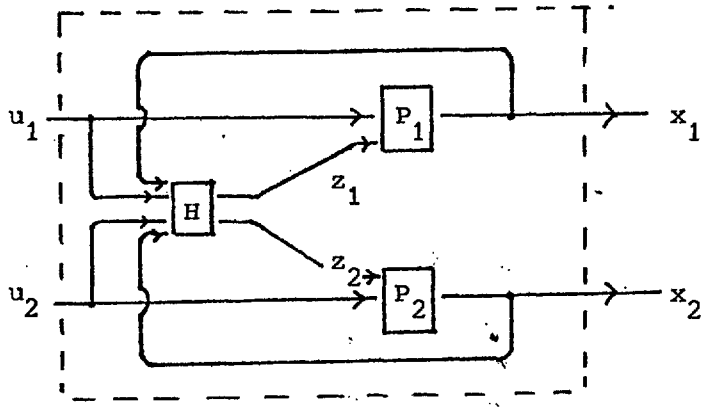


fig 2

In fig 2 the dotted lines enclose the global process $P : U \to X$. Its internal structure is designated as two sub-processes P_1 and P_2, with inputs u_i, z_i and outputs x_i, the interactions arising from the coupling mode H with inputs u_i, x_i.

The local decision problems can now be formulated in terms of an objective function $\bar{J}_i$ given on $U_1 \times Z_i$, the i^{th}, subprocess, and a performance function $\bar{G}_i : U_i \times Z_i \times X_i \to V$

$$\bar{J}_i\ (u_i,z_i) = \bar{G}_i[u_i,z_i,P_i(u_i,z_i)]. \tag{10}$$

§3. Coordination principles

Following [39] we consider the following basic coordination principles, by which a designated "coordinator" may influence the subproblems with a view to achieving a solution of the global problem. Two cases arise :

Case 1: Model coordination

Choose a coordination input $\alpha = (\alpha_1 \ldots \alpha_N)$ in the coordination space Z, and consider the subprocess model $P_{i\alpha}(u_i) = P_i(u_i,\alpha_i)$. Then for each $\alpha \in Z$ the local control problem $\mathcal{D}_i(\alpha)$ is to *find a control* $\hat{u}_i$ *in* U_i such that

$$\bar{J}_i(\hat{u}_i,\alpha_i) = \min_{U_i} \bar{J}_i(u_i,\alpha_i) \tag{11}$$

Case 2: Goal coordination

Choose a parameter $\beta \in B$ such that a performance function $G_{i\beta} : U_i \times Z_i \times X_i \to V$ is a modification of the original $\bar{G}_i$. Let the cost function $\bar{J}_i$ be defined on $U_i \times Z_i$ in terms of P_i and $G_{i\beta}$ as in (10). Then for each $\beta \in B$ the local control problem $\mathcal{D}_i(\beta)$ is to *find a pair* $(\hat{u}_i,\hat{z}_i)$ *in* $U_i \times Z_i$ s.t.

$$\bar{J}_{i\beta}(\hat{u}_i,\hat{z}_i) = \min_{U_i \times Z_i} \bar{J}_{i\beta}(u_i,z_i). \tag{12}$$

Note that Case 1 involves minimization with respect to u only, but Case 2 with respect to u and z, i.e. in Case 2 the interaction inputs are treated as free variables. In other words the coupling constraint no longer operates, and the system is fictitiously "decoupled". We shall see that this leads to problems of non-feasible controls computed iteratively.

We now postulate a coordination principle for each of these cases. We shall then investigate under what conditions the principle may be applied.

Case 1: Interaction prediction principle (for model coordination)

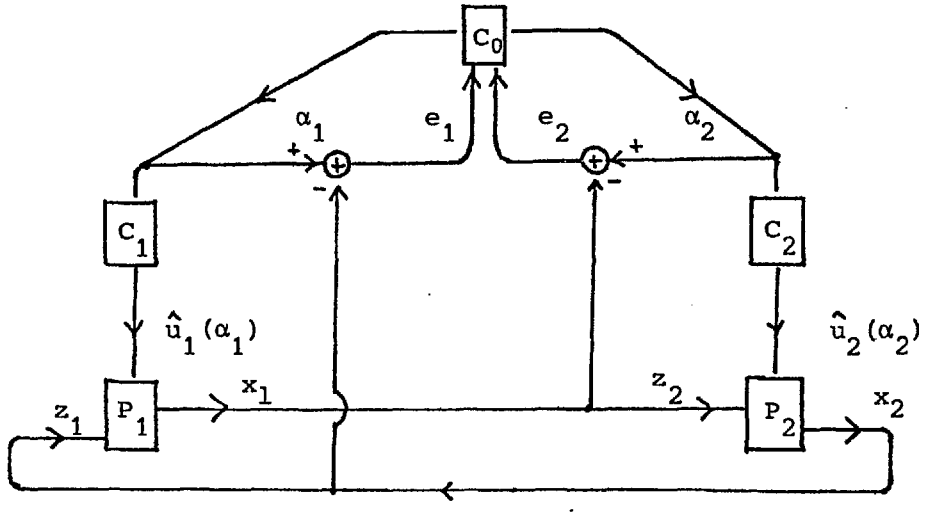
Let $\alpha = (\alpha_1 \ldots \alpha_N)$ be the predicted coordination inputs, and let $z_1(\alpha), \ldots z_N(\alpha)$ be the actual coordination inputs occurring when the local optimal control $\hat{u}(\alpha)$ is implemented. The interaction prediction principle then asserts that the *overall* optimum is achieved if the predicted coordination inputs are correct, i.e.

$$\alpha_i = z_i(\alpha) \qquad i = 1 \ldots N \tag{13}$$

This principle is not necessarily valid for a particular problem. When it is valid, the supremal controller makes a prediction α^k of the coordination inputs, then measures the error e^k between α^k and the actual inputs $z^k(\alpha^k)$, then corrects or updates the original prediction α^k to α^{k+1}.

If the IP principle holds, it yields the supremal coordination problem $\mathcal{P}_0$: *find* $\hat{\alpha}$ *in* Z *s.t.* $e = \| \hat{\alpha} - z(\hat{\alpha}) \|$ *is minimized.*

The IP principle is illustrated in fig. 3, in the context of a 2-level, 2 subsystem process, where $z_1 = x_2, z_2 = x_1$.



Case 2: Interaction balance principle (for goal coordination)

Let $\beta \in B$ be given; let $\hat{z}(\beta)$ be the coordination inputs required by the local controllers to achieve local optima ; let $z(\beta)$ be the coordination inputs that occur if the optimal control $\hat{u}(\beta)$ is implemented, i.e. $z(\beta) = K(\hat{u}(\beta))$. The interaction balance principle then states that the *overall optimum* is achieved if the actual coordination inputs precisely equal those required for local optimization, i.e.

$$\hat{z}_i(\beta) = z_i(\beta) \quad , \quad i = 1,\ldots N \tag{14}$$

The supremal controller chooses a value β^k of the parameter β, and measures the error between the demanded coordination inputs $\hat{z}(\beta)$ and the actual inputs arising from application of local optimal controls $\hat{u}(\beta)$, then corrects the parameter β^k to some β^{k+1}.

If the IB principle holds, the supremal control problem $\mathcal{D}_0$ is : find β in B s.t. $e = \| \hat{z}(\beta) - z(\beta) \|$ is minimized.

The principle is illustrated in fig 4.

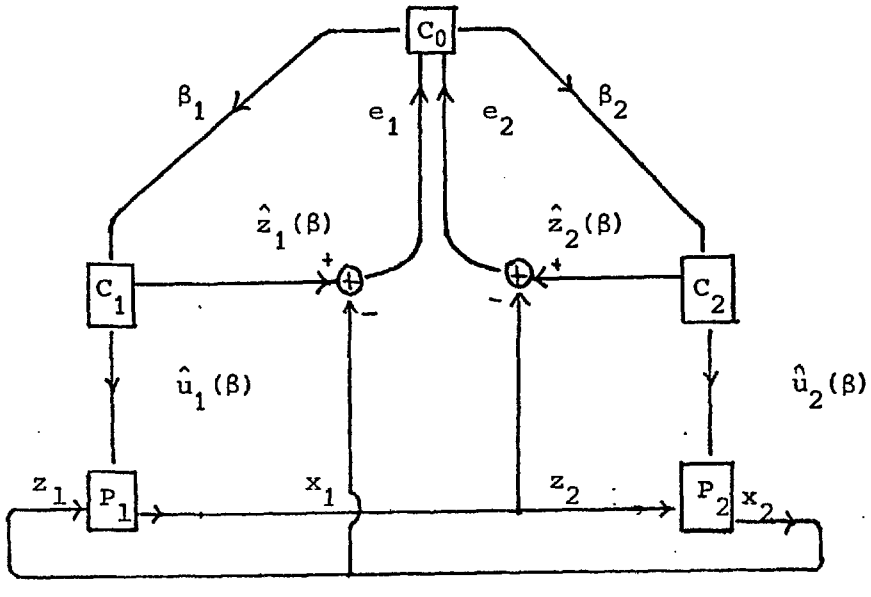


fig 4

It may be seen that the IP and IB principles are based on the interaction constraint

$$z = H(u,x).$$

It is important to realize however that satisfaction of this constraint is neither a necessary nor a sufficient condition for attaining a global optimum, *in general*. Hence the necessity to investigate which systems are coordinable by these principles, and under what conditions they are applicable. We illustrate the problem in the following example.

Example 1 (Bernhard [7])

Let U,X,V be $\mathbb{R}^2$, and define the global process

$$\dot{x} = P(u)$$

by

$$x_1 = 2u_1 + 2u_2$$
$$x_2 = u_1 + 3u_2,$$

and the performance $G(u,x)$ to be minimized by

$$G(u,x) = 2u_1^2 + (x_1-7)^2 + u_2^2 + (x_2+7)^2.$$

Let the subprocesses $x_i = \bar{P}_i(u_i,z_i)$ be defined by

$$x_1 = 2u_1 + 2z_1$$
$$x_2 = 3u_2 + z_2 \qquad (15)$$

and the coupling function $z = K(u)$ by

$$z_1 = u_2$$
$$z_2 = u_1 \quad . \qquad (16)$$

We choose local performances

$$G_1(u_1,x_1) \triangleq 2u_1^2 + (x_1-7)^2$$
$$G_2(u_2,x_2) \triangleq u_2^2 + (x_2+7)^2, \qquad (17)$$

so that $G_1 + G_2 = G$.

Substituting for x in G(u,x) gives

$$J(u) = 2u_1^2 + (2u_1 + 2u_2 - 7)^2 + u_2^2 + (u_1 + 3u_2 + 7)^2$$
$$= 7(u_1^2 + 2u_2^2) + 2u_1u_2 - 2u_1 + 2u_2 + 14)$$

and it is easily shown that $u^* = (3,-2)$, when

$$J(u^*) = 63.$$

Let us apply the IP principle to this problem.

(a) If we take the prediction $\alpha = (-\frac{7}{2}, \frac{14}{3})$, we have by(15) that

$$x_1 = 2u_1 - 7$$
$$x_2 = 3u_2 + \frac{14}{3}$$

and substituting in (17) we have that

$$G_1(u_1,x_1) = J_1(u_1) = 2u_1^2 + (2u_1 - 14)^2$$
$$G_2(u_2,x_2) = J_2(u_2) = u_2^2 + (3u_2 + \frac{35}{3})^2 \quad ,$$

so that

$$\hat{u} = (\frac{14}{3}, -\frac{7}{2}).$$

Hence by (16), $z = K(\hat{u})$, i.e.

$$z = (-\frac{7}{2}, \frac{14}{3}) = \alpha$$

We see that the coupling is respected, although the locally optimal control $\hat{u}$ is not the global optimum u^*. Nor do $\hat{u}$ and u^* generate the same cost, for $G(\tilde{u},P(\hat{u}))$ is $72\frac{5}{18}$.

Therefore we see that satisfaction of the coupling function, or interconnection constraint (5), is not a sufficient condition for attaining the global optimum.

(b) Let us take the prediction $\alpha = (-1, -\frac{1}{3})$. Then by (15) we have

$$x_1 = 2u_1 - 2$$
$$x_2 = 3u_2 - \frac{1}{3}$$

and substituting in (17) we have that

$$G_1(u_1,x_1) = J_1(u_1) = 2u_1^2 + (2u_1-9)^2$$

$$G_2(u_2,x_2) = J_2(u_2) = u_2^2 + (3u_2 + \frac{20}{3})^2$$

so that

$$\hat{u} = (3,-2) = u^*$$

but $\alpha \neq K(\hat{u})$, for by (16)

$$K(\hat{u}) = (-2,3) \neq (-1,-\frac{1}{3}) \quad .$$

Hence satisfaction of the coupling constraint is not a necessary condition for attaining the global optimum. □

This is an illuminating example, shedding light on the problem of how one knows when a coordination procedure has attained the global optimum. A natural test is to check whether the necessary conditions of global optimality are met. But this usually requires extensive system-wide calculation, which is not in the spirit of decentralization, and it does not offer a means of adjusting the parameter α if the test is not satisfied. If we direct our attention to satisfaction of the coupling constraint, we see from the above example that this is neither a necessary nor a sufficient condition for attaining the global optimum. This motivates the following definitions.

Definition

We say that the IP principle is *applicable* if

$$\alpha = K(\hat{u}) \Rightarrow \hat{u} = u^* \tag{18}$$

and the IB principle is *applicable* if,

$$\hat{z}(\beta) = K(\hat{u}) \Rightarrow \hat{u} = u^* \; . \tag{19}$$

The implication (18) is a restatement of the IP test (13);

Similarly (19) is a restatement of (14). In the light of these definitions, we have the task of identifying classes of problems to which these principles are *applicable*.

Example 1 (continued)

(c) Let us check whether an α satisfying (18) exists. If it exists, it must be $\alpha = K(u^*)$, i.e. from (16),

$$\alpha = (-2,3).$$

But if we apply this prediction, we have from (15) that

$$x_1 = 2u_1 - 4$$

$$x_2 = 3u_2 + 3$$

and substituting in (17) gives

$$G_1(u_1) = 2u_1^2 + (2u_1 - 11)^2$$
$$G_2(u_2) = u_2^2 + (3u_2 + 10)^2$$

so that

$$\hat{u} = \left(\frac{11}{3}, -3\right) \neq u^*, \text{ and } G(\hat{u}) = 70\frac{7}{9} \geq 63 = G(u^*).$$

Hence no such α exists. This motivates the next definition.

Definition

We say that a system is coordinable by the IP principle if the principle is applicable, and if there exists α^* s.t. the local problems have a solution $\hat{u}(\alpha^*)$, and that

$$\alpha^* = K(\hat{u}(\alpha^*)).$$

Similarly we say that a system is *coordinable* by the IB principle if the principle is applicable, and if there exists β^* such that the local problem have a solution $\hat{u}(\beta^*)$ and that

$$\hat{z}(\beta^*) = K[\hat{u}(\beta^*)] \qquad (21)$$

Let us now consider an example of IB principle (interaction decoupling).

Example 2 (Mesarovic [39])

$$x_1 = 2u_1 + z_1$$
$$x_2 = 2u_2 - z_2 \tag{22}$$
$$G_i(u_i, x_i) = u_i^2 + (x_i - 1)^2 \tag{23}$$
$$z_1 = x_2$$
$$z_2 = x_1 \tag{24}$$

We modify the G_i as follows :

$$\begin{aligned} \bar{G}_{1\beta}(u_1, z_1, x_1) &= G_1 + \beta_1 z_1^2 - \beta_2 x_1^2 \\ \bar{G}_{2\beta}(u_2, z_2, x_2) &= G_2 + \beta_2 z_2^2 - \beta_1 x_2^2 . \end{aligned} \tag{25}$$

We note that modifications can be called *zero-sum* since, substituting (24), we have that

$$\begin{aligned} &\beta_1 z_1^2 - \beta_2 x_1^2 + \beta_2 z_2^2 - \beta_1 x_2^2 \\ &= \beta_1 x_2^2 - \beta_2 x_1^2 + \beta_2 x_1^2 - \beta_1 x_2^2 \\ &= 0 . \end{aligned}$$

Substituting (22) in (25), we have the modified local criteria :

$$\bar{J}_{1\beta}(u_1, z_1) = u_1^2 + (2u_1 + z_1 - 1)^2 + \beta_1 z_1^2 - \beta_2 (2u_1 + z_1)^2$$

$$\bar{J}_{2\beta}(u_2, z_2) = u_2^2 + (2u_2 - z_2 - 1)^2 + \beta_2 z_2^2 - \beta_1 (2u_2 - z_2)^2 .$$

The optimal $\hat{u}, \hat{z}$ then satisfy the following necessary conditions:

$$\begin{bmatrix} 5-4\beta_2 & 2-2\beta_2 \\ 2-2\beta_2 & 1+\beta_1-\beta_2 \end{bmatrix} \begin{bmatrix} u_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 5-4\beta_1 & -2+2\beta_1 \\ -2+2\beta_1 & 1-\beta_1+\beta_2 \end{bmatrix} \begin{bmatrix} u_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

subject to the matrices being positive definite (to ensure a minimum).

Take $\beta = (0,-\frac{1}{2})$, then it is readily found that

$$\begin{pmatrix} \hat{u}_1 \\ \hat{z}_1 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{2}{3} \end{pmatrix} \quad ; \quad \begin{pmatrix} \hat{u}_2 \\ \hat{z}_2 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \end{pmatrix}$$

and indeed $\hat{u} = (0, \frac{2}{3}) = u^*$. Hence this problem is coordinable by the IB principle.

§4. Interaction operators

We now consider how the infimal performance measures may be modified by what Mesarovic et al call 'interaction operators' to ensure applicability of the coordination principles. We concentrate on the IP principle, and leave discussion of the IB principle to the section on *dual methods* below [§5].

Recall that our basic model is

$$x = P(u), \tag{26}$$

or in 'decoupled' form

$$x = \bar{P}(u,z) \tag{27}$$

where $$z = H(u,x) \tag{28}$$

or $$z = K(u) = H(u,P(u)).$$

The aggregate performance function is $G(u,x)$, with variants

$$J(u) = G(u,P(u)) \tag{29}$$

and $$\bar{J}(u,z) = G(u,\bar{P}(u,z)) \tag{30}$$

The overall problem is to minimize $J(.)$ in U. As stated by Cohen [6], 'the purpose of decentralization is to define, for each $i = 1,\ldots N$, functionals ψ_i^0 over U_i such that by independent minimization of each ψ_i^0, each optimum u_i^k is equal to the i^{th} part u_i^* of u^*, the global optimum. Since it is not possible, generally,

to achieve this directly, an iterative process, called "coordination", is required to transform the ψ_i^0 's into ψ_i^1 's, ..., ψ_i^k 's etc, such that the respective optima u_i^1 's,..., u_i^{k+1} 's etc, converge toward the u_i^* 's.

We construct suitable local criteria ψ_i^0 by considering the effect of a perturbation δu in the control. This perturbation exerts a direct effect on the overall system cost, but also an indirect effect via the interactions propagated throughout the system by the coupling function. These direct and indirect effects can be studied by considering the 'decentralized' version $\bar{J}(u,z)$ of the system cost.

Let the current control be u^0, so that $x^0 = P(u^0)$ and $z^0 = K(u^0)$. Current cost is $J(u^0)$ or $\bar{J}(u^0, K(u^0))$. (31)

Now consider a perturbation δu_i in the i^{th} component u_i^0 of u^0, making this component $u_i = u_i^0 + \delta u_i$, so that

$$u^0 + \delta u_i = (u_i^0, \ldots u_{i-1}^0, u_i^0 + \delta u_i, u_{i+1}^0, \ldots u_N^0) \tag{32}$$

In the 1970 text, Mesarovic, Macko and Takahara [20]employ the notation $u^0[u_i]$ for $u^0 + \delta u_i$, while in his 1977 paper, Cohen [6] employs the notation $u^0 - u_i^0 + u_i$, where $u_i = (0, \ldots 0, u_i, 0, \ldots 0)$ [not to confuse it with an element of U_i], and in this case,

$$u_i = u_i^0 + \delta u_i.$$

As a result of this perturbation, the new cost is $J(u^0 + \delta u_i)$ or

$$\bar{J}(u_i^0 + \delta u_i, K(u^0 + \delta u_i)). \tag{33}$$

If we neglect the interaction effects, and concentrate only on the *direct* effects of the perturbation, the new cost is

$$\bar{J}(u^0 + \delta u_i, K(u^0)), \tag{34}$$

while if we neglect the direct effects, and concentrate only on the indirect effects of the perturbation carried by the interactions, the new cost is

$$\bar{J}(u^0,\ K(u^0 + \delta u_i)). \tag{35}$$

Mesarovic defines the difference between (35) and (31) as the total goal interaction, which in turn defines the goal-interaction *operator* $\gamma: U \times U \to V$, by

$$\gamma(u^0,\ \delta u_i) = \bar{J}(u^0, K(u^0 + \delta u_i)) - \bar{J}(u^0, K(u^0)).$$

Other operators expressing the sensitivity of other versions of the cost and other system parameters to a perturbation in control, may be defined, but they will not concern us here.

Now we consider first-order linear approximations to these quantities by taking differentials. To first order, we have that, overall,

$$\delta J \simeq < \frac{dJ}{du}(u^0),\ \delta u> \tag{37}$$

where δJ is given exactly by

$$\delta J = J(u^0+\delta u) - J(u^0) . \tag{38}$$

Again direct and indirect effects are captured using the 'decentralized' version $\bar{J}(u,z)$:

$$\delta J \simeq < \frac{\partial \bar{J}}{\partial u}(u^0, K(u^0)),\ \delta u> + < \frac{dz}{du}(u^0)^T \frac{\partial \bar{J}}{\partial z}(u^0, K(u^0)), \delta u> \tag{39}$$

where the right hand side of (39) is the matrix form of the familiar scalar chain rule :

$$\frac{dJ}{du} = \frac{\partial \bar{J}}{\partial u} + \frac{\partial \bar{J}}{\partial z}\ \frac{dz}{du} . \tag{40}$$

The direct effects of δu are contributed via the first term on the right hand side of (39), and the indirect effects through the second term. This second term is thus a linear approximation to Mesarovic's total goal interaction, and so allows us to define a *linearized* goal-interaction operator $\Gamma: U \times U \to V$, by

$$\Gamma(u^0)\delta u = < \frac{dz}{du}(u^0)^T\ \frac{\partial \bar{J}}{\partial z}(u^0, K(u^0)),\ \delta u> \tag{41}$$

where the operator itself is defined in matrix terms :

$$\Gamma(u^0) = \left(\frac{dz}{du}(u^0)^T \ \frac{\partial \bar{J}}{\partial z}(u, K(u^0))\right)^T. \tag{42}$$

We now show that if we define the infimal performance functionals ψ_i^0 by 'restricting' the local costs to take account of direct effects only (i.e. neglecting the interaction effects propagated through the coupling function) and then 'modifying' them by the linearized goal-interaction operator, we make the problem coordinable by the IP principle.

Following Cohen [6], we define $\psi_i^0(u_i)$ as

$$\psi_i^0(u_i) = \bar{J}(u^0 + u_i - u_i^0, K(u^0)) + \Gamma_i(u^0)\, u_i \tag{43}$$

This definition is justified by the following lemmas. We assume all functionals are strictly convex and differentiable in u, and in fact that interconnections are linear (to ensure decomposibility), and injective.

with modification by interaction operator.

Lemma 1 (Applicability of IP principle) (Cohen)

If for some interaction prediction $\alpha = K(u^0)$ the resulting interconnection $z = K(u')$, where u' is the solution to

$$\frac{\partial \psi_i^0}{\partial u_i}(u_i') = 0, \qquad i = 1 \ldots N, \tag{44}$$

is such that

$$\alpha = z,$$

then $u^0 = u' = u^*$.

Proof Since K is injective, we have that

$$\alpha = z \Rightarrow u^0 = u' \tag{45}$$

Now from (43),

$$\frac{\partial \psi_i^0}{\partial u_i}(u_i) = \frac{\partial \bar{J}}{\partial u_i}(u^0 + u_i - u_i^0, K(u^0)) + \Gamma_i(u^0)$$

Aggregating, we have from (44) that

$$0 = \frac{\partial \psi^0}{\partial u}(u') = \frac{\partial \bar{J}}{\partial u}(u', K(u^0)) + \Gamma(u^0)$$

$$= \frac{\partial \bar{J}}{\partial u}(u^0, K(u^0)) + \Gamma(u^0) \quad \text{by (45)}$$

$$= \frac{dJ}{du}(u^0) \quad \text{by (39) and (42).}$$

i.e. $u^0 = u'$ satisfies optimality condition $\frac{dJ}{du}(u') = 0$. □

Note that Lemma 1 is stated in exactly the form required for *applicability* of the IP principle, as spelt out in (18) above. Note too that injectivity of K is a severe requirement that considerably restricts the scope of this principle. This feature is not usually noted in the literature.

Next we check coordinability, as defined by (20).

Lemma 2 (Coordinability of IP principle) (Cohen) with modification by interaction operator

There exists a prediction $\alpha^0 = K(u^0)$ s.t. the local problems have a solution u' and s.t.

$$\alpha^0 = K(u').$$

Proof

We argue by contradiction. Let us assume that for all predictions $\alpha = K(u)$ the local problems have a solution u' such that

$$\alpha \neq K(u') \tag{46}$$

In particular, take $\alpha^* = K(u^*)$ where u^* is the global optimum. Again, by injectivity of K, (46) implies that

$$u^* \neq u' \tag{47}$$

Now by definition of u^* we have that

$$0 = \frac{dJ}{du}(u^*) = \frac{\partial \bar{J}}{\partial u}(u^*, K(u^*)) + \Gamma(u^*) \tag{48}$$

and by definition of u', obtained from u^*,

$$0 = \frac{\partial \psi^0}{\partial u}(u') = \frac{\partial \bar{J}}{\partial u}(u', K(u^*)) + \Gamma(u^*). \tag{49}$$

But by convexity of J, there is at most one $\bar{u}$ such that

$$\frac{dJ}{du}(\bar{u}) = 0,$$

namely u^*, and hence there is at most one $\bar{u}$ such that

$$\frac{\partial \bar{J}}{\partial u}(\bar{u}, K(u^*)) + \Gamma(u^*) = 0,$$

namely, u'. Hence

$$u^* = u' = \bar{u} ,$$

which contradicts (47). This completes the proof. □

We have shown that the the modified system is such that the IP principle is applicable (Lemma 1) and satisfies condition (20); hence it is coordinable.

We could state general theorems specifying what conditions on systems guarantee that the IP (and IB) principles succeed, but this would take us too far afield - suffice to note that the LQ regulator can certainly be modified to allow the IP principle to work. We shall show explicitly how this can be done in the case of the static regulator in Chapter 3.

The coordination algorithm is only completely specified when the procedure for updating ψ_i^k is given. For instance, if 'pure reinjection' or successive approximations is used, the local optimum u^{k+1} obtained from a given u^k is used as the next prediction, hence defining ψ^{k+1} through $\Gamma(u^{k+1})$. The complete algorithm would then be : given a current control u^k, and hence $x^k = P(u^k)$ and $z^k = K(u^k)$, the modified and restricted local problems are:

$$\text{find} \quad u_{k+1} \in \arg\min_{u_i} \{\bar{J}_i(u_i, K(u^k)) + \Gamma_i(u^k)u_i\} \tag{48}$$

$$\left.\begin{aligned} \text{i.e., find} \quad u_{k+1} \in \arg\min_{u_i} & \; G_i(u_i, x_i) + \Gamma_i(u^k)u_i \\ \text{s.t.} \quad x_i &= \bar{P}_i(u_i, z_i) \\ z_i &= K_i(u^k) \; . \end{aligned}\right\} \tag{49}$$

Of course the convergence of such a scheme has not been settled - all we have done is specify a principle and found a problem on which it can work. No general results on the convergence of the scheme given are known. We study convergence in the concrete instance of the LQ regulator in Chapter 3 and 5 below.

We illustrate the foregoing notions on Example 1 above.

Example 1 (contd)

Recall that $x = P(u)$ given by $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ (50)

while $x = \bar{P}(u,z)$ is $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$

..... (51)

and $z = K(u)$ is $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ (52)

The overall performance $G(u,x)$ is $2u_1^2 + (x_1-7)^2 + u_2^2 + (x_2+7)^2$ (53)

which decomposes into

$$G_1(u_1,x_1) = 2u_1 + (x_1-7)^2$$
$$G_2(u_2,x_2) = u_2^2 + (x_2+7)^2 \qquad (54)$$

Substituting (50) into (53) gives

$$J(u) = G(u,P(u)) = 7\{u_1^2 + 2u_2^2 + 2u_1u_2 - 2u_1 + 2u_2 + 14\}$$

..... (55)

and substituting (51) gives

$$\bar{J}(u,z) = G(u,\bar{P}(u,z)) = 6u_1^2 + 4z_1^2 + 8u_1z_1 - 28u_1 - 28z_1 + 10u_2^2 + z_2^2 + 6u_2z_2 + 42u_2 + 14z_2 + 98$$

..... (56)

Computing gradients, we have that

$$\frac{dJ}{du} = \left(\frac{\partial J}{\partial u_1}, \frac{\partial J}{\partial u_2} \right)^T = 7(2u_1 + 2u_2 - 2,\ 4u_2 + 2u_1 + 2)^T \qquad (57)$$

$$\frac{\partial \bar{J}}{\partial u} = \left(\frac{\partial \bar{J}}{\partial u_1}, \frac{\partial \bar{J}}{\partial u_2} \right)^T = 2(6u_1 + 4z_1 - 14,\ 10u_2 + 3z_2 + 21)^T \qquad (58)$$

$$\frac{\partial \bar{J}}{\partial z} = \left(\frac{\partial J}{\partial z_1}, \frac{\partial \bar{J}}{\partial z_2} \right)^T = 2(4u_1 + 4z_1 - 14,\ 3u_2 + z_2 + 7)^T \qquad (59)$$

and

$$\frac{dz}{du} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} . \qquad (60)$$

Take u^0 as $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Then $x^0 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$ by (50), and $z^0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ by (52). $G(u^0, x^0) = 133$, and it is easily checked that $J(u^0) = 133$ and $\bar{J}(u^0, z^0) = 133$.

Now consider a perturbation in u_1 from 1 to 2. Hence $u = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, and now $x = \begin{bmatrix} 6 \\ 5 \end{bmatrix}$ and $z = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Note that $\delta u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\delta x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\delta z = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. It is now found that $G(u,x) = 154$, so $\delta J = 21$.

What are the direct and indirect effects of this perturbation? Taking direct effects only, we have that $\bar{J}(u, K(u^0)) = \bar{J}\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) = 131$, while indirect effects taken alone give

$$\bar{J}(u^0, K(u)) = \bar{J}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right) = 156.$$

What are the first order approximations in this case? We have that

$$\frac{dJ}{du}(u^0) = \begin{bmatrix} 14 \\ 56 \end{bmatrix}$$

$$\frac{\partial \bar{J}}{\partial u}(u^0) = \begin{bmatrix} -8 \\ 68 \end{bmatrix}$$

and

$$\frac{\partial \bar{J}}{\partial z}(u^0) = \begin{bmatrix} -12 \\ 22 \end{bmatrix} .$$

Note in passing that $\frac{\partial \bar{J}}{\partial u} + K^T \frac{\partial \bar{J}}{\partial z} = \begin{pmatrix} -8 \\ 62 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -12 \\ 22 \end{pmatrix} = \begin{pmatrix} 14 \\ 56 \end{pmatrix} = \frac{dJ}{du}$

First order approximation to cost change is

$$\delta J = \left\langle \frac{dJ}{du}(u^0), \delta u \right\rangle = (14 \quad 56) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 14.$$

The interaction operator in this case is

$$\Gamma(u^0)^T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -12 \\ 22 \end{pmatrix} = \begin{pmatrix} 22 \\ -12 \end{pmatrix}$$

Hence $\psi^0(\delta u) = \bar{J}(u, K(u^0)) + \langle \Gamma(u^0), \delta u \rangle$

$$= 131 + (22 \quad -12) \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$= 153$$

which is very close to $J(u) = 154$.

§5. Primal and dual techniques

Primal and dual strategies are particular realizations of the interaction prediction and interaction balance principles. They are the oldest of the decomposition procedures, dating back to the early work of Pearson, Lasdon and others in the 1960's. A number of different formulations of these procedures have appeared in the literature. To discuss these we consider the primal system

$$\text{(P)} \qquad \begin{aligned} x_i &= \bar{P}_i(u_i, z_i) \\ z_i &= H_i(u, x) \end{aligned} \qquad i = 1, \ldots, N \tag{53}$$

and the criteria to be minimized

$$\bar{G}(u,x,z) = \sum_{i=1}^{N} G_i(u_i, x_i, z_i) \tag{54}$$

Frequently the dual problem is posed in terms of the Lagrange dual with respect to the interconnection constraint $z = H(u,x)$:

$$\text{(D)} \qquad \max_{\mu} \min_{u,x,z} L(x,u,z,\mu)$$

$$\text{subject to } x = \bar{P}(u,z), \tag{55}$$

where

$$L(x,u,z,\mu) = \bar{G}(x,u,z) + \mu^T(z-H(u,x)). \tag{56}$$

It may be seen that this is a particular instance of the IB principle, where the coordinating variable β is given by the Lagrange dual variable μ.

The validity of this procedure derives from the Lagrange Multiplier Theorem for convex constrained optimization problems :

$$\min_{x\in X} f(x) \quad \text{s.t.} \quad \phi(x) = 0 \tag{57}$$

where $f : X \to R$ is a convex functional on a normed linear space, and $\phi: X \to Z$, which takes values in the normed linear space Z, is continuously differentiable.

We say that x_0 is a regular point for ϕ if $\phi'(x)$ maps X onto Z, where ' denotes the derivative. In the case of finite dimensional maps $\phi : \mathbb{R}^n \to \mathbb{R}^m$, a point $X_0 \in \mathbb{R}^n$ is a regular point for ϕ if the Jacobian matrix of ϕ has rank m.

The Lagrange multiplier theorem now states that if f achieves a local extremum at x_0 which is a regular point for ϕ, then there exists $\lambda_0^* \in X^*$ such that the Lagrangian functional

$$L(x,\lambda^*) = f(x) + \lambda_0^*\phi(x) \tag{50}$$

is stationary at x_0, i.e.

$$f'(x_0) + \lambda_0^* \phi'(x_0) = 0 \tag{59}$$

An equivalent formulation is that there exists $\lambda_0^* \in X^*$ such that the Lagrangian has a saddle point at (x_0,λ_0^*), i.e.

$$\left\{ \begin{array}{l} f(x_0) \le f(x) \quad \forall x \\ \text{and } \phi(x_0) = 0 \end{array} \right\} \Longleftrightarrow \{L(x_0,\lambda) \le L(x_0,\lambda_0^*) \le L(x,\lambda_0^*)\}.$$

Hence we say the Lagrange dual to problem (57) is

$$\max_{\lambda} \min_{x} \; f(x) + \lambda^{*} \phi(x) \tag{60}$$

which is how we arrived at (55).

We shall formulate both primal and dual approaches in a general form that allows us to deduce a number of coordination procedures. For simplicity of exposition we are assuming that (P) is a finite dimensional mathematical programming problem.

First let us form the Lagrangian with respect to both interconnection and system constraints:

$$\bar{L}(x,u,z,\lambda,\mu) = G(x,u,z) + \lambda^{T}(x-\bar{P}(u,z)) + \mu^{T}(z-H(u,x)) \tag{61}$$

We note that if the interconnections are *linear* :

$$z_i = \sum_{\substack{j=1 \\ j\neq i}}^{N} H^{1}_{ij} u_j + \sum_{\substack{j=1 \\ j\neq i}}^{N} H^{2}_{ij} x_j \tag{62}$$

then the Lagrangian can be decomposed into N sub-Lagrangians

$$\bar{L}_i(x_i,u_i,z_i,\lambda_i,\mu_i) = \bar{G}_i(x_i,u_i,z_i) + \lambda_i^{T}(x_i-\bar{P}_i(u_i,z_i)) + \mu_i^{T} z_i$$
$$- x_i^{T} \sum_{\substack{j=1 \\ j\neq i}}^{N} H^{1T}_{ji} \mu_j - u_i^{T} \sum_{\substack{j=1 \\ j\neq i}}^{N} H^{2T}_{ji} \mu_j \tag{63}$$

where we use the identity

$$\sum_{i=1}^{N} \mu_i^{T} \sum_{\substack{j=1 \\ j\neq i}}^{N} H_{ij} v_j \equiv \sum_{i=1}^{N} v_i^{T} \sum_{\substack{j=1 \\ j\neq i}}^{N} H_{ji}^{T} \mu_j \tag{64}$$

It is this possibility of decomposition that underlies all the 'parallel computation' procedures that have been suggested.

If we assume that the system constraints are independent, then we have the following set of necessary conditions :

$$\left.\begin{aligned}
&\frac{\partial L_i}{\partial x_i} = 0 && \text{1}\\
&\frac{\partial L_i}{\partial u_i} = 0 && \text{11}\\
&\qquad\qquad i=1,\ldots,N\\
&\frac{\partial L_i}{\partial z_i} = 0 && \text{111}\\
&\frac{\partial L_i}{\partial \lambda_i} = 0 && \text{iv}\\
&\frac{\partial L_i}{\partial \mu_i} = 0 && \text{v}
\end{aligned}\right\} \qquad (65)$$

Satisfaction of (65) (i) - (v) i = 1,...,N, simultaneously gives a full set of optimal values $(x^*, u^*, z, \lambda^*, \mu^*)$. Satisfaction of a subset of them, and iterative resetting of the complementary variables at a supremal level, leads to a number of possible multilevel schemes. Such schemes would only make sense if

(i) they lead to parallel computation at the infimal level;

(ii) the supremal update problem is simple;

(iii) information transmission requirements are not excessive.

Let us call the coordinating variable v, and the complementary or *local* variables, w. Then the pattern of information flow between the two levels is

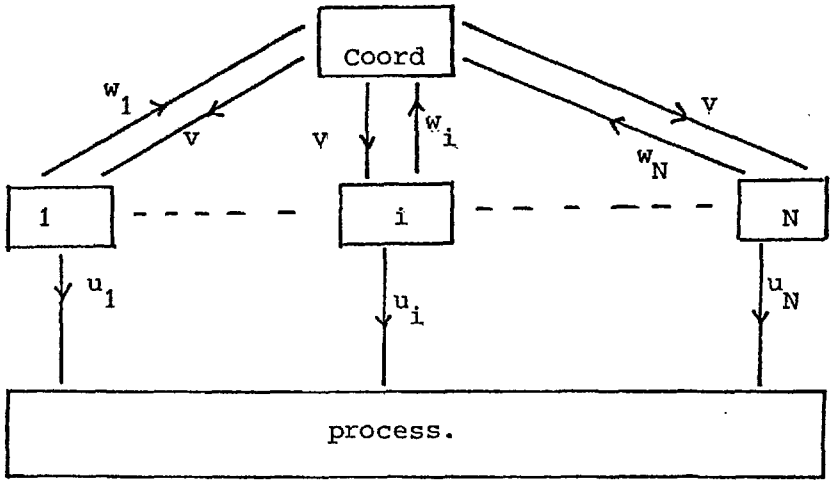


fig.5

Let us take as coordinating variable v the system state x. Then this would be useful if the N problems

$$\left.\begin{array}{l} \dfrac{\partial L_i}{\partial u_i}(x) \\ \dfrac{\partial L_i}{\partial z_i}(x) \\ \dfrac{\partial L_i}{\partial \lambda_i}(x) \\ \dfrac{\partial L_i}{\partial \mu_i}(x) \end{array}\right\} = 0 \quad , \quad i = 1,\ldots,N \tag{66}$$

had a solution that could be implemented in decentralized parallel fashion. The supremal updating task could then be based on the gradient of the Lagrangian with respect to x:

$$\nabla_x L = \frac{\partial L}{\partial x} + \frac{\partial L}{\partial u}\frac{du}{dx} + \frac{\partial L}{\partial z}\frac{dz}{dx} + \frac{\partial L}{\partial \lambda}\frac{d\lambda}{dx} + \frac{\partial L}{d\mu}\frac{d\mu}{dx} \tag{67}$$

If at the infimal level, $\frac{\partial L}{\partial u} = \frac{\partial L}{\partial z} = \frac{\partial L}{\partial \lambda} = \frac{\partial L}{\partial \mu} = 0$, then it is sufficient to take

$$\nabla_x L = \frac{\partial L}{\partial x} \tag{68}$$

The 'dual procedure' has usually been interpreted strictly to mean the use of μ as a coordinating variable. The primal procedure has variously been interpreted to mean the use of x,z or a combination of them as coordinating variable. It turns out (see examples taken in Chapter 3) that strict use of x,z leads to non-solvable infimal problems a point not noted, for example, in Simmons and White [11] . Other formulations of the primal approach, e.g. Findeisen[15], Tatjewski [19], do not specify how the "primal variables" are to be selected, and so they avoid such problems. We review this formulation.

In Findeisen [15] the technique of using desired values of sub-

system outputs as coordination instruments is described. The technique is called a direct (or primal, parametric, or feasible) mode of coordination.

As before the system is described in decentralized fashion as

$$x = \bar{P}(u,z) \ , \tag{69}$$

$$z = H(u,x) \tag{70}$$

where, for Findeisen, dependence of interaction terms on controls is ignored, and (70) is written as the linear relation

$$z = Hx$$

$$\text{or} \qquad z_i = \sum_{j=1}^{N} H_{ij}x_j = H_i x. \tag{71}$$

With each subsystem a performance functional $\bar{J}_i(u_i,z_i)$ is associated, and again

$$J(u) = \bar{J}(u,HP(u)) = \sum_{i=1}^{N} \bar{J}_i \tag{72}$$

A coordinator chooses a coordinating variable v, giving local problems :

$$\min_{u_i \in U_i(v)} \bar{J}_i(u_i,H_i v) \tag{73}$$

where $$U_i(v) = \{u_i \,|\, v_i = \bar{P}_i(u_i,H_i v)\} \tag{74}$$

and the coordination variable v *prescribes* the subsystem outputs (and hence interaction inputs, via the coupling function).

The goal of the coordinator is to minimize the overall performance function :

$$\min_{u \in V_* \cap V_o} \bar{J}(\hat{u}(v),\ HP(\hat{u}(v))) \tag{75}$$

where $$V_* = \{v \in X : P(\hat{u}(v)) \in X\} \neq \chi \tag{76}$$

$$V_o = \{v \in X : U_i(v) \neq \phi, \ \forall i = 1,\ldots,N\} \tag{77}$$

and $\hat{u}(v) = (\hat{u}_1(v), \ldots, \hat{u}_N(v))$, assuming that $\hat{u}(v)$ exists for every v in $V_* \cap V_o$.

The structure of the direct coordination procedure is as shown in fig. 6.

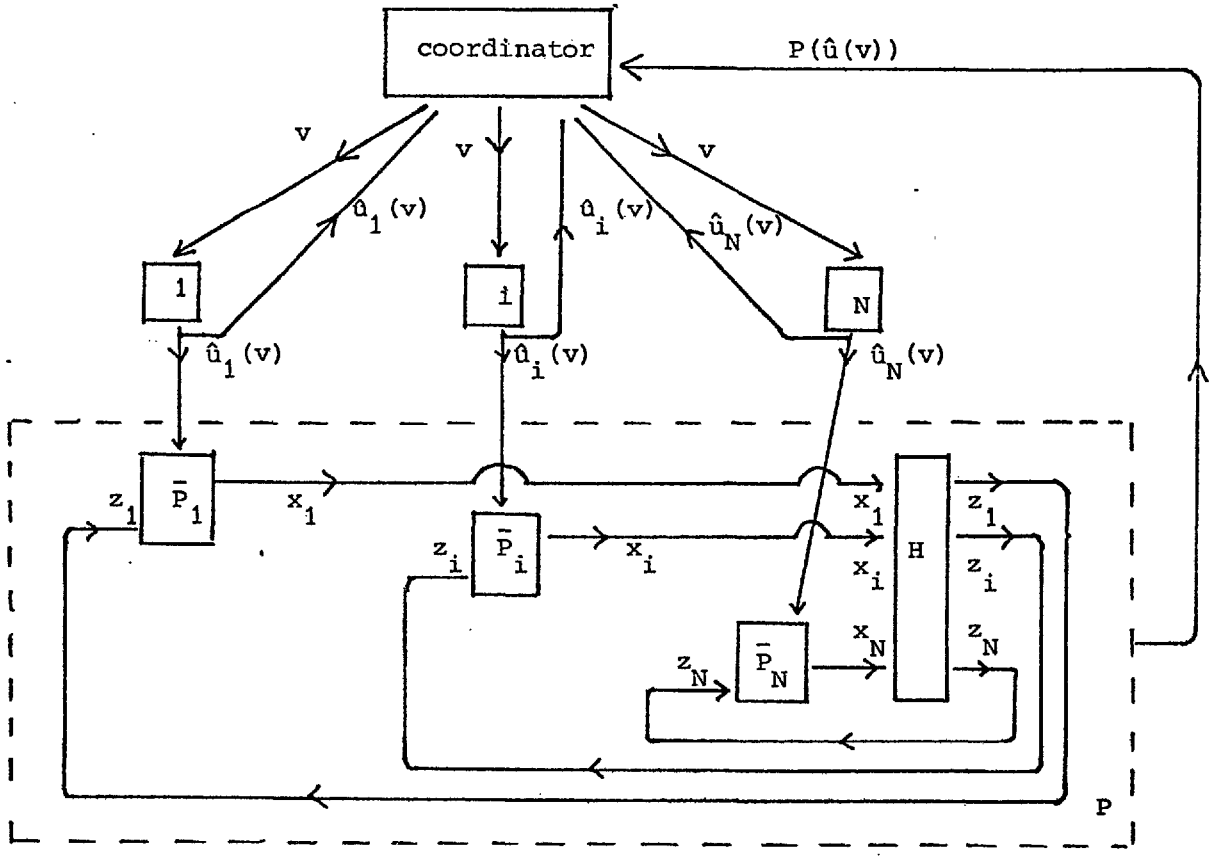


fig. 6

Remark 1

The coordinator uses subsystem outputs as decision variables. The local decision units solve typical off-line optimization problems based on subsystem models. The coordinator is not

allowed to set the controls u_i directly, i.e. it cannot over-ride local decisions.

Remark 2

The method does not specify how the coordinating variable is to be chosen, giving a degree of flexibility in this matter to the coordinator.

Remark 3

The feasible set V_o is not specified explicitly in the initial problem. Except for special cases, it is difficult, if at all possible, to determine the feasible set. This is a drawback of the primal method.

Remark 4

We may note that although there appear to be no applicability or coordinability conditions attached to the use of this method, they are in fact buried in the feasible set. For if $v \epsilon V_o$, there must exist $u_i \epsilon U_i(v)$ (coordinability) s.t. $v_i = \bar{P}_i(u_i, z_i)$ where $z_i = H_i v$ (applicability). Findeisen notes the conditions under which these conditions are met : all system maps P, $\bar{P}_i$, H_i, $\bar{J}_i$, i,...,N, continuous, and unique solutions at infimal and supremal levels (guaranteed by strict convexity).

Finally we should note that we have not discussed the iteration nor convergence of these primal and dual schemes. Again this is a subject that would take us too far afield from our present concerns. Suffice to note that the standard convexity assumptions on the cost, and linear assumptions on the interconnections, certainly guarantee convergence of both primal and dual schemes. These assumptions will always be met by our LQ regulator.

§6 Unified approach to decomposition schemes

In 1978 Cohen [25] published a unified approach to the problem of optimization by decomposition and coordination. Although we do not use Cohen's formulation in the sequal, it is nevertheless an important innovation, and sheds interesting light on the above procedures. Cohen built his unified approach from two basic principles : the auxiliary problem principle, and the relaxation principle. Combining them, he obtained a general algorithm which could be specialized to many particular cases. We shall indicate how it may be specialized to the case of Findeisen's primal scheme presented in §5 above.

The basic optimization problem is

$$\min_{u \in U} J(u) \tag{78}$$

where J is a continuous, convex and differentiable functional defined on a Hilbert space U.

The Auxiliary Problem (AP) principle derives from the auxiliary functional

$$G^v(u) = K(u) + \langle \varepsilon J'(v) - K'(v), u \rangle \tag{79}$$

where ' denotes the derivative, and where K is a functional, called the core, which has the same characteristics as J. We then have the AP principle:

Lemma 3 (Cohen)

If v minimizes $G^v(.)$, then it also minimizes $J(.)$

Outline of proof

Since K is differentiable, G^v is differentiable; hence

$$\frac{dG^v}{du}(v) = K'(v) + \varepsilon J'(v) - K'(v) = \varepsilon J'(v). \tag{80}$$

But if v minimizes $G^v(.)$, then

$$\frac{dG^v}{du}(v) = 0$$

Hence by (80)

$$J'(v) = 0 \text{ and so } v \text{ minimizes } J . \qquad \square$$

The AP principle generates an algorithm if we set the current control u^k to be v. Hence we have

AP algorithm 1

(i) Choose u^o. Set $k = 0$.

(ii) Solve the AP^k :

$$\min_{u \in U} \; K(u) + \langle \varepsilon J'(u^k) - K'(u^k), u \rangle$$

and let u^{k+1} be a solution

(iii) Stop if stopping condition satisfied ; otherwise set $k = k+1$ and go to (ii).

We now give two examples to show that appropriate choice of the core functional K yields well-known algorithms.

Example 3 (Cohen) $K(u) = \frac{1}{2}\|u\|^2$ (81)

In this case

$$G^v(u) = \tfrac{1}{2}\|u\|^2 + \langle \varepsilon J'(v) - v, u \rangle$$

so that (82)

$$\frac{dG^v}{du}(u) = u + \varepsilon J'(v) - v$$

and equating to zero gives

$$u^{k+1} = u^k - \varepsilon J'(u^k) \qquad (83)$$

by setting $u = u^{k+1}$ and $v = u^k$ in (82). Thus we obtain the steepest descent algorithm.

Example 4 (Cohen) $K(u) = \frac{1}{2}\langle u, J''(v)u \rangle$. (84)

In this case

$$G^v(u) = \tfrac{1}{2}\langle u, J''(v)u \rangle + \langle \varepsilon J'(v) - J''(v)v, u \rangle$$

so that

$$\frac{dG^v}{du}(u) = J''(v)u + \varepsilon J'(v) - J''(v)v \qquad (85)$$

and equating to zero gives

$$u^{k+1} = u^k - \varepsilon J''(u^k)^{-1} J'(u^k) \tag{86}$$

where again we set $u^{k+1} = u$ and $u^k = v$ in (85).

Thus we obtain the Newton-Raphson algorithm.

The AP principle is subsequently related to parallel decomposition (although as seen above there is no intrinsic connection). The Relaxation principle is now introduced, related to sequential decompositions.

Let $U = U_1 x \ldots x U_N$.

Then we have the well-known

Relaxation algorithm 2

(i) Choose $u^o = (u_1^o, \ldots u_N^o)$. Set $t = 0$, $i=1$.

(ii) Solve

$$\min_{u_i \in U_i} J(u_1^{t+1}, \ldots u_{i-1}^{t+1}, u_i, u_{i+1}^t, \ldots u_N^t) \tag{87}$$

and let u_i^{t+1} be a solution.

(iii) If $i = N$ go to (iv), otherwise set $i = i+1$ and return to (iii).

(iv) Stop if stopping condition satisfied; otherwise set $i=1$, $t=t+1$, and return to (ii). □

We now combine the principles of relaxation and the AP into one algorithm. First we reformulate the basic problem as

$$\min_{(u,v) \in U x V} J(u,v) \tag{88}$$

where both U and V can be decomposed.

This will be solved by successively minimizing an augmented function over U and then V iteratively.

Hence we have the

Relaxed AP algorithm 3

(i) Choose (u^o, v^o). Set k=0

(ii) Solve the AP_u^k :

$$\min_{u \in U} K(u, v^k) + \langle \varepsilon_1 J_u'(u^k, v^k) - K_u'(u^k, v^k), u \rangle \qquad (89)$$

and let u^{k+1} be a solution.

(ii) Solve the AP_v^k :

$$\min_{v \in V} K(u^{k+1}, v) + \langle \varepsilon_2 J_v'(u^{k+1}, v^k) - K_v'(u^{k+1}, v^k), v \rangle$$

and let v^{k+1} be a solution.

(iv) Stop if stopping condition satisfied; otherwise set k=k+1 and return to (ii). □

We now make the connection between this relaxed AP algorithm 3 and decomposition schemes by specifying that AP_v^k constitutes a coordination task, while AP_u^k breaks up into N parallel minimizations at the infimal level.

By making the following specifications we may, for example, recover Findeisen's primal scheme from algorithm 3. Let us set

$$J = \sum_{i=1}^{N} J_i \qquad (91)$$

$$\varepsilon_1 = 1 \qquad (92)$$

$$K(u, v) = \sum_{i=1}^{N} K_i(u_i) + \tfrac{1}{2}\|v\|^2 \qquad (93)$$

$$K_i(u_i) = J((u_1^k, \ldots u_{i-1}^k, u_i, u_{i+1}^k, \ldots u_N^k), v^k) \qquad (94)$$

where (u^k, v^k) is given.

Then the infimal problem (89) is

$$\min_{u \in U} \sum J_i((u_1^k, \ldots u_{i-1}^k, u_i^k, u_{i+1}^k, \ldots, u_N^k), v^k) + \langle J_u'(u^k, v^k), u \rangle - \langle J_u'(u^k, v^k), u \rangle. \qquad (95)$$

.....

(ignoring the terms in v only) and this splits into N problems

$$\min_{u_i \in U_i} J_i((u_1^k, \ldots u_{i-1}^k, u_i, u_{i+1}^k, \ldots u_N^k), v^k). \tag{96}$$

The supremal problem (90) gives the steepest descent algorithm, as in Example 3 above:

$$v^{k+1} = v^k - \epsilon_2 J_v'(u^{k+1}, v^k). \tag{97}$$

The combination of (96) and (97) is Findeisen's scheme with a particular technique of updating the coordination variable v.

CHAPTER 3

The steady state interconnected linear quadratic regulator

§1. Introduction and outline of chapter

§2. Formulation of model and derivation of optimality conditions

§3. Multilevel schemes derived from Lagrangian Theory

§4. Supremal coordination schemes: decomposition based on successive approximations

§5. Gradient methods and interaction prediction

§6. Relations betweeen gradient and Lagrangian schemes

§1. Introduction

In this chapter we apply the general schemes outlined in Chapter 2 to the concrete instance of the steady state LQ regulator. There is already a considerable literature on "multilevel mathematical programming", and the influential Polish School of Findeisen has derived a number of primal, or parametric, schemes in the context of steady-state control schemes. Nevertheless a systematic investigation of decomposition /coordination schemes applied to the steady-state LQ regulator seems not to have been attempted before.

The plan of the chapter is as follows. We first set up the model in §2 and derive the necessary conditions for optimal behaviour. In §3 we systematically attack the Lagrange necessary conditions to investigate the possibilities of 2-level decomposition procedures. It turns out that some schemes formulated in the abstract in previous published papers fail in this most simple of concrete instances - hence the utility of a systematic investigation. Supremal updating schemes are assumed to be gradient-based, and it is shown how the gradient of the system Lagrangian is easily found. In §4 an alternative approach is introduced, and the supremal updating scheme is assumed to be of 'reinjection' type. This is generalized as the application of successive approximations (i.e. Jacobi iteration) to the linear Lagrange conditions. From this generalized approach the steady-state analogue of the well-known Takahara scheme is derived, together with a range of other schemes which are not derivable from the multilevel schemes discussed in §3, and which have never before been described. A general convergence result is given. Finally the question of speeding up convergence of these schemes by use of Gauss-Seidel iteration is discussed, and the

surprising observation is made that the Takahara scheme is already in Gauss-Seidel form.

In §5 an entirely different perspective is adopted, and in place of Lagrange multipliers, the system *gradient* is made the focus of attention. This task of computing the current gradient (with respect to the current control) is assigned to a supremal coordinator, and the problem is then to find gradient-based algorithms which permit parallel computation by a number of substations supplied with partial information on the gradient. This line of attack, initiated by Cohen, yields a number of decomposition schemes, such as the conjugate gradient (K^{-1}) scheme and Newton Raphson $(K)^{-1}$ scheme, where K is chosen to give suitable decomposition. This latter scheme is then shown to be identical to the interaction prediction approach using Mesarovic's interaction operator. The identity has previously been established indirectly by Cohen, but here it is shown directly in terms of control sequences.

Finally in §6 the relations between the Lagrangian and gradient approaches are discussed. Steady state analogues of the dynamic costate equation are derived. It is shown in detail how Cohen's characterization of the Lagrangian-based Takahara scheme as a local approximation to the interaction prediction scheme, can be justified. Other possible relations are mentioned, and left as open questions.

§2. Formulation of model and derivation of optimality conditions

We define a large-scale linear system S as an interconnection of N subsystems $S_1, \ldots S_N$ by the linear equation

$$A_i x_i + B_i u_i + z_i = 0 \tag{1}$$

where $x_i \in \mathbb{R}^{n_i}$ is a vector of intermediate outputs (states) of S_i ,

$u_i \in \mathbb{R}^{m_i}$ is a vector of control variables for S_i ,

$Z_i \in \mathbb{R}^{n_i}$ is a vector of intermediate inputs to S_i .

Intermediate inputs, controls and outputs are related linearly by an interconnection constraint of the form

$$Z_i = \sum_{j=1}^{N} G_{ij} u_j + \sum_{j=1}^{N} H_{ij} x_j \quad , \quad i = 1 \ldots N \tag{2}$$

In terms of system aggregates

$$x = (x_1, x_2, \ldots x_N)^T \text{ , etc} \tag{3}$$

we may write the overall system equation in a connected form as

$$Ax + Bu + z = 0 \tag{4}$$

$$z = Gu + Hx \tag{5}$$

or in unconnected form as

$$(A + H)x + (B + G)u = 0 \ , \tag{6}$$

where $A \in \mathbb{R}^{nxn}$, $B \in \mathbb{R}^{nxm}$ are assumed to be block-diagonal, and

$H \in \mathbb{R}^{nxn}$ and $G \in \mathbb{R}^{nxm}$ are the system interconnection matrices, defined in terms of H_{ij}, G_{ij}, $i = 1 \ldots N$, $j = 1 \ldots N$, where $G_{ii} = H_{ii} = 0$.

A typical example of such a system would be a chemical plant

where a number of subunits are interconnected, and system variables represent flows of real products. A schematic representation of the set up is given in fig 1.

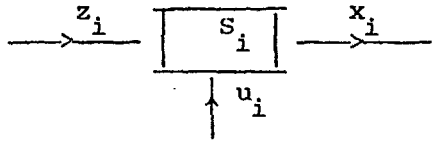


fig 3.1

To introduce the notion of optimal performance of this system, we associate a cost with each subsystem of the form

$$\bar{G}_i(u_i, x_i, z_i) = \tfrac{1}{2}x_i^T Q_i x_i + \tfrac{1}{2}u_i^T R_i u_i + \tfrac{1}{2}z_i^T S_i z_i - b_i^T x_i \qquad (7)$$

where $Q_i = Q_i^T \geq 0$ is positive semi definite,

and $R_i = R_i^T > 0$ is positive definite,

and $S_i = S_i^T > 0$ is positive definite.

The addition of the term in z_i is certainly non-standard in quadratic studies, and amounts (as we see below) to including cross-coupling terms between u and x in the overall system cost. These terms are needed in one of the decomposition procedures - otherwise we can ignore the term in z_i, and set $S_i = 0$.

If we assume that overall costs are additive in subsystem costs, i.e.

$$\bar{G} = \Sigma \bar{G}_i$$

then the regulator can be written in aggregate form as

$$\min \bar{G}(u,x,z) = \tfrac{1}{2}(x^T Q x + u^T R u + z^T S z) - b^T x \qquad (8)$$

$$\text{s.t.} \quad 0 = Ax + Bu + z$$

$$z = Gu + Hx$$

Eliminating z, we have that

$$G(u,x) = \tfrac{1}{2}\{x^T(Q+H^TSH)x + u^T(R+G^TSG)u + x^TH^TSGu + u^TG^TSHx\} - b^Tx)$$

..... (9)

$$\text{s.t.} \quad 0 = (A+H)x + (B+G)u \ ,$$

or if $S = 0$,

$$G(u,x) = \tfrac{1}{2}x^TQx + \tfrac{1}{2}u^TRu - b^Tx \ . \tag{10}$$

We assume that A^{-1} and $(A+H)^{-1}$ exists, so that we may solve (6) and (4) as

$$x = -(A+H)^{-1}(B+G)u \tag{11}$$

and $$\dot{x} = -A^{-1}Bu - A^{-1}z. \tag{12}$$

This may appear to be a strong assumption, but it is in fact made implicitly by all models which assert a relation of the form

$$x = P(u)$$

or $$x = \bar{P}(u,z) \tag{13}$$

as done, for instance, by Mesarovic [19], Simmons and White [11] and others. For subsequent use we may derive a number of interrelated forms of the regulator. Using (5) and (11) we have (still using Mesarovic formalism)

$$z = Gu - H(A+H)^{-1}(B+G)u = K(u) \tag{14}$$

so that

$$\begin{aligned}
\bar{P}(u,K(u)) &= -A^{-1}Bu - A^{-1}Gu + A^{-1}H(A+H)^{-1}(B+G)u \\
&= [A^{-1}H(A+H)^{-1} - A^{-1}]\ (B+G)u \\
&= A^{-1}[H(A+H)^{-1} - (A+H)(A+H)^{-1}](B+G)u \\
&= A^{-1}[H - (A+H)](A+H)^{-1}(B+G)u \\
&= -(A+H)^{-1}(B+G)u \\
&= P(u).
\end{aligned}$$

We may also derive expressions for $J(u) = G(u,P(u))$ and $\bar{J}(u,z) = G(u,\bar{P}(u,z))$ but it is more convenient to do this in §5 on gradient schemes.

Let us now derive optimality conditions using Lagrangian methods. Assuming $S = 0$ for the moment, we have the basic problem.

$$\left.\begin{aligned} &\min_u \; G(u,x) = \tfrac{1}{2}x^TQx + \tfrac{1}{2}u^TRu - b^Tx \\ &\qquad\qquad \text{s.t.} \quad (A+H)x + (B+G)u = 0 \end{aligned}\right\} \tag{15}$$

The Lagrangian for this problem is

$$L(x,u,\lambda) = \tfrac{1}{2}x^TQx + \tfrac{1}{2}u^TRu - b^Tx + \lambda^T[(A+H)x + (B+G)u] \tag{16}$$

and the necessary conditions for optimality are

$$\left.\begin{aligned} &\frac{\partial L}{\partial \lambda} = (A+H)x + (B+G)u = 0 \\ &\frac{\partial L}{\partial x} = Qx - b + (A+H)^T\lambda = 0 \\ &\frac{\partial L}{\partial u} = Ru + (B+G)^T\lambda = 0 \,. \end{aligned}\right\} \tag{17}$$

Hence $\hat{u} = -R^{-1}(B+G)^T\lambda$

where

$$\begin{bmatrix} A+H & -(B+G)R^{-1}(B+G)^T \\ Q & (A+H)^T \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix} . \tag{18}$$

We shall return to this global set of necessary conditions in the context of successive approximations coordination strategies.

Now let us employ Lagrangian techniques in tackling the interconnected problem

$$\left.\begin{aligned} &\min_u \; \tfrac{1}{2}x^TQx + \tfrac{1}{2}u^TRu + \tfrac{1}{2}z^TSz - b^Tx \\ &\qquad\qquad \text{s.t.} \quad Ax + Bu + z = 0 \\ &\qquad\qquad\qquad\quad z = Gu + Hx. \end{aligned}\right\} \tag{19}$$

In this case the Lagrangian is given by

$$L(x,u,\lambda,\mu) = \tfrac{1}{2} x^T Q x + \tfrac{1}{2} u^T R u + \tfrac{1}{2} z^T S z - b^T x + \lambda^T (Ax + Bu + z) + \mu^T (z - Gu - Hx) \quad \ldots\ldots \quad (20)$$

and the necessary conditons for optimality are:

$$\begin{aligned}
\frac{\partial L}{\partial \lambda} &= Ax + Bu + z = 0 \\
\frac{\partial L}{\partial x} &= Qx + A^T\lambda - H^T\mu - b = 0 \\
\frac{\partial L}{\partial u} &= Ru + B^T\lambda - G^T\mu = 0 \\
\frac{\partial L}{\partial z} &= Sz + \lambda + \mu = 0 \\
\frac{\partial L}{\partial \mu} &= z - Gu - Hx = 0
\end{aligned}$$

or

$$\begin{bmatrix} A & 0 & B & I & 0 \\ Q & A^T & 0 & 0 & -H^T \\ 0 & B^T & R & 0 & -G^T \\ 0 & I & 0 & S & I \\ -H & 0 & -G & I & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \\ u \\ z \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \ldots\ldots \quad (21)$$

Let us check first that (21) is in fact the same as the global necessary conditions (18). Setting $S = 0$, we have that

$$\mu = -\lambda$$

$$z = Gu + Hx$$

$$u = -R^{-1}(B+G)^T\lambda$$

so that

$$Ax - BR^{-1}(B+G)^T\lambda - GR^{-1}(B+G)^T\lambda + Hx = 0$$

$$Qx + A^T\lambda + H^T\lambda = b$$

i.e.

$$\begin{bmatrix} A+H & -(B+G)R^{-1}(B+G)^T \\ Q & (A+H)^T \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix} \quad (21a)$$

which is (18) again. Hence (21) represents a genuinely more structured version of (18), that is to say it contains more terms, but on elimination of terms it reduces to the same expression.

§3. Multilevel schemes derived from Lagrangian theory

In the context of this simple model it is of interest to classify all possible coordination schemes, to see which lead to viable decomposition schemes, and which do not. Such an exercise has not been published before, although the spirit of the approach has been paralleled by Mahmoud, Vogt and Mickle [56] using what they call a 'generalized gradients' technique, and by that adopted by Singh and Titli [8].

We begin with a general formulation that will enable us to derive all subsequent schemes.

Let (x,u,z) be aggregrated in a vector h. Then (19) may be reformulated as

$$\min_{h} G(h)$$

$$\text{s.t.} \quad Z(h) = b \ .$$

Various combinations of local and coordinating variables v,w may be given as

$$h = Mv + Nw,$$

and different arrangements of the system and interconnection constraints as

$$P[Z(h)-b] = 0$$

and $$(I-P)[Z(h)-b] = 0 .$$

Here P and I-P are projection operators. Now all types of coordination using the Lagrangian approach may be formulated as

$$\max_{p} \left\{ \min_{v} \left[\begin{array}{c} \min_{w} \; G(Mv + Nw) + \langle p, P[Z(Mv+Nw)-b] \rangle \\ \text{s.t. } (I-P)[Z(Mv+Nw) - b] = 0 \end{array} \right] \right\} . \qquad (22)$$

(1) Dual interaction coordination [goal coordination]

In this scheme the interaction multiplier μ is set at the supremal level and used as coordination variable. The infimal problem is then to solve (21) apart from $\frac{\partial L}{\partial \mu} = 0$, i.e. (21) (i) - (iv) :

given $\bar{\mu}$,

$$Ax + Bu + z = 0 \qquad \text{(i)}$$

$$Qx + A^T\lambda - H^T\bar{\mu} = b \qquad \text{(ii)}$$

$$Ru + B^T\lambda - G^T\bar{\mu} = 0 \qquad \text{(iii)}$$

$$Sz + \lambda + \bar{\mu} = 0 \; . \qquad \text{(iv)}$$

The scheme may be derived from the general formulation (22) by setting $p = \mu$,

$$P[Z(h)-b] = z - Gu - Hx \; ,$$

$$(I-P)[Z(h)-b] = Ax + Bu + z = 0.$$

Solving (21) (i) - (iv), we have from (iii),

$$u = -R^{-1}B^T\lambda + R^{-1}G^T\bar{\mu}$$

and from (iv),

$$z = -S^{-1}\lambda - S^{-1}\bar{\mu}.$$

Substituting for u and z then gives

$$Ax - (BR^{-1}B^T + S^{-1})\lambda = -(BR^{-1}G^T - S^{-1})\bar{\mu}$$

$$Qx + A^T\lambda = H^T\bar{\mu} + b$$

or

$$\begin{bmatrix} A & -BR^{-1}B^T - S^{-1} \\ Q & A^T \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} -BR^{-1}G^T + S^{-1} \\ H^T \end{bmatrix} \bar{\mu} + \begin{bmatrix} 0 \\ b \end{bmatrix} . \qquad (23)$$

We see that (23) may be decomposed as

$$\begin{pmatrix} A_i & -B_i R_i^{-1} B_i^T - S_i^{-1} \\ Q_i & A_i^T \end{pmatrix} \begin{pmatrix} x_i \\ \lambda_i \end{pmatrix} = \begin{pmatrix} -B_i R_i^{-1} \sum_{j=1}^{N} G_{ji}^T \bar{\mu}_j + S_i^{-1} \bar{\mu}_i \\ \sum_{j=1}^{N} H_{ji}^T \bar{\mu}_j + b_i \end{pmatrix} \quad (24)$$

for i = 1 ... N. Hence it can be solved in a decentralized fashion, provided the matrix

$$\begin{pmatrix} A_i & -B_i R_i^{-1} B_i^T - S_i^{-1} \\ Q_i & A_i \end{pmatrix}$$

is invertible.

Alternatively we may consider the following infimal problem directly :

$$\min_{u_i} \sum_{i=1}^{N} \{G_i(u_i, x_i, z_i) + \bar{\mu}_i^T [z_i - \sum_{j=1}^{N} G_{ij} u_j - \sum_{j=1}^{N} H_{ij} x_j]\}. \quad (25)$$

Using the identity

$$\sum_{i=1}^{N} \bar{\mu}_i^T \sum_{j=1}^{N} H_{ij} x_j = \sum_{j=1}^{N} x_i^T \sum_{j=1}^{N} H_{ji}^T \bar{\mu}_j \quad , \quad \text{ie } \mu^T H x = x^T H^T \mu$$

we may rewrite (25) as

$$\min \sum_{i=1}^{N} \{G_i(u_i, x_i, z_i) + \bar{\mu}_i^T z_i - u_i^T \sum_{j=1}^{N} G_{ji}^T \bar{\mu}_j - x_i^T \sum_{j=1}^{N} H_{ji}^T \bar{\mu}_j\}$$

.... (26)

s.t. $\quad 0 = A_i x_i + B_i u_i + z_i.$

Adjoining the system constraint in (26) we obtain a Lagrangian which splits into N sub-Lagrangians

$$L_i(x_i, u_i z_i \lambda_i, \bar{\mu}_i) = G_i(u_i, x_i, z_i) + \lambda_i^T(A_i x_i + B_i u_i + z_i) + \mu_i^T z_i - u_i^T \sum_{j=1}^{N} G_{ji}^T \bar{\mu}_j - x_i^T \sum_{j=1} H_{ji}^T \bar{\mu}_j \qquad (27)$$

and the necessary conditons for optimality

$$\left.\begin{aligned} \frac{\partial L_i}{\partial \lambda_i} &= A_i x_i + B_i u_i + z_i = 0 \\ \frac{\partial L_i}{\partial x_i} &= Q_i x_i + A_i^T \lambda_i - \sum H_{ji}^T \bar{\mu}_j = b_i \\ \frac{\partial L_i}{\partial u_i} &= R_i u_i + B_i^T \lambda_i - \sum G_{ji}^T \bar{\mu}_j = 0 \\ \frac{\partial L_i}{\partial z_i} &= S_i z_i + \lambda_i + \bar{\mu}_i = 0 \end{aligned}\right\} \qquad (28)$$

which is seen to be the decomposed form of (21).

Hence $u_i = -R_i^{-1} B_i^T \lambda_i + R_i^{-1} \sum G_{ji}^T \bar{\mu}_j$

and $z_i = -S_i^{-1} \lambda_i - S_i^{-1} \bar{\mu}_i$

and substitution in (28) gives (24).

Because the local criterion in (27) is considerably modified, this method has been called goal modification.

Supremal coordination

Let the solution of the subproblems obtained from (28) be $u_i^*(\bar{\mu})$, $x_i^*(\bar{\mu})$, $\lambda_i^*(\bar{\mu})$ and $z_i^*(\bar{\mu})$, $i = 1 \ldots N$. Now let us employ a gradient technique to update μ at the supremal level. Recall that the gradient of the system Lagrangian with respect to μ is

$$\nabla_{\mu} L = \left(\frac{\partial L}{\partial x}\right)^T \frac{dx}{d\mu} + \left(\frac{\partial L}{\partial u}\right)^T \frac{du}{d\mu} + \left(\frac{\partial L}{\partial z}\right)^T \frac{dz}{\partial \mu} + \left(\frac{\partial L}{\partial \lambda}\right)^T \frac{d\lambda}{d\mu} + \left(\frac{\partial L}{\partial \mu}\right)^T$$

and since we set $\frac{\partial L}{\partial x} = \frac{\partial L}{\partial u} = \frac{\partial L}{\partial z} = \frac{\partial L}{\partial \lambda} = 0$,

we have that

$$\nabla_{\mu} L^{T} = \frac{\partial L}{\partial \mu} = z^{*} - Gu^{*} - Hx^{*}. \tag{29}$$

Hence μ could be updated at supremal level by the steepest ascent algorithm :

$$\mu^{k+1} = \mu^{k} + \varepsilon(z^{k} - Gu^{k} - Hx^{k}) . \tag{30}$$

Note that the coupling constraint

$$z = Gu + Hx.$$

is not satisfied until the algorithm converges; hence this is a non-feasible technique. That is to say it is a 'fictitious' technique that employs a 'decoupled' system in which interconnections are not satisfied. The information required is as shown in fig. 2.

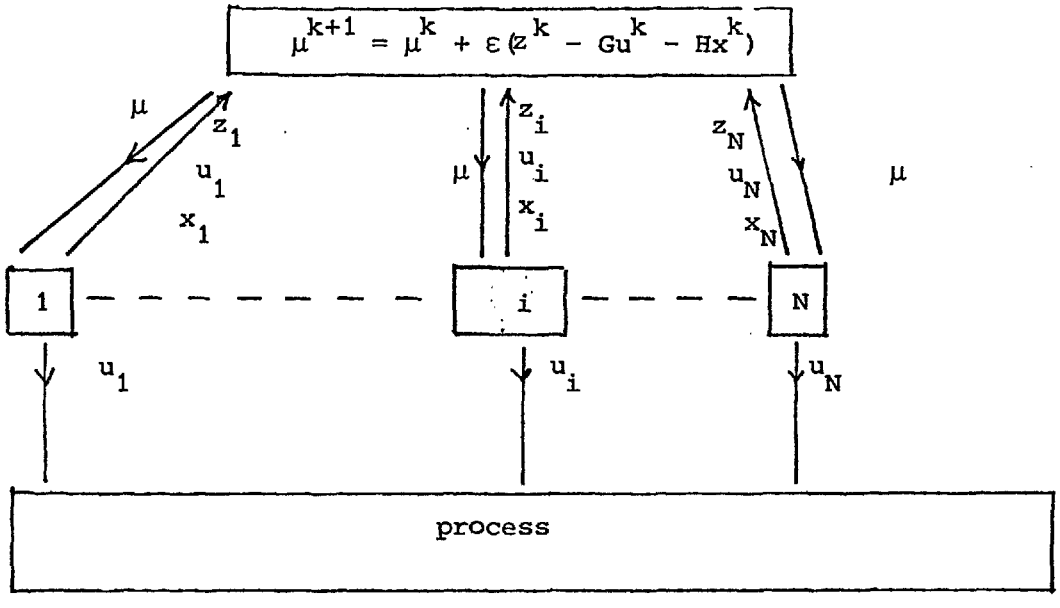


fig. 3.2

(2) Other schemes with a single vector being updated at supremal level.

In [56], Mahmoud, Vogt and Mickle propose using λ as coordination variable [a scheme they dub as Co-Co, or Costate Coordination]; in [8] Singh, Titli propose using x as coordination variable [a scheme they call 'model coordination'] while in [56] again, Mahmoud et al propose using z as coordination variable, and *they* call this 'model coordination'. Noone has proposed using u, but it is equally a candidate. A quick inspection of (21) shows why none of these schemes in fact yields a decomposition procedure : they all leave $\frac{\partial L}{\partial \mu} = z - Gu - Hx$ to be solved at the infimal level. It is worth spelling out just why this vitiates decomposition.

(i) z as coordination variable

In this case the infimal problem is to solve (21) apart from $\frac{\partial L}{\partial z} = 0$, i.e. (i), (ii), (iii) and (v) :

$$\text{given } \bar{z}, \quad \begin{aligned} Ax + Bu + \bar{z} &= 0 \\ Qx + A^T\lambda - H^T\mu &= b \\ Ru + B^T\lambda - G^T\mu &= 0 \\ \bar{z} - Gu - Hx &= 0 \end{aligned} \tag{31}$$

In decomposed form we have

$$\begin{aligned} A_i x_i + B_i u_i + \bar{z}_i &= 0 \\ Q_i x_i + A_i^T \lambda_i - \sum H_{ji}^T \mu_j &= b_i \\ R_i u_i + B_i^T \lambda_i - \sum G_{ji}^T \mu_j &= 0 \\ \bar{z}_i - \sum G_{ij} u_j - \sum H_{ij} x_j &= 0 \ . \end{aligned}$$

From (iii) we have that

$$u_i = -R_i^{-1} B_i^T \lambda_i + R_i^{-1} \sum G_{ji}^T \mu_j \ ,$$

and from (ii) (if $Q > 0$)

$$x_i = -Q_i^{-1} A_i^T \lambda_i + Q_i^{-1} \sum H_{ji}^T \mu_j + Q_i^{-1} b_i ,$$

and substituting in (i) and (v) :

$$-(A_i Q_i^{-1} A_i^T + B_i R_i^{-1} B_i^T)\lambda_i + (A_i Q_i^{-1} \sum H_{ji}^T + B_i R_i^{-1} \sum G_{ji}^T)\mu_j = -\bar{z}_i - A_i Q_i^{-1} b_i$$

$$(\sum G_{ij} R_j^{-1} B_j^T + \sum H_{ij} Q_j^{-1} A_j^T)\lambda_j - (\sum H_{ij} Q_j^{-1} (\sum H_{ji}^T) + \sum G_{ij} R_j^{-1} (\sum G_{ji}^T))\mu_j$$

$$= -\bar{z}_i + Q_i^{-1} b_i$$

It is clear that taking z as coordinating variable requires each subsystem to know values of other subsystem parameters - hence the idea of decomposition is non-viable.

(ii) x as coordination variable

The infimal problem is to solve (21) apart from $\frac{\partial L}{\partial x} = 0$, i.e. (i), (iii), (iv) and (v), given $\bar{x}$:

$$\begin{aligned} A\bar{x} + Bu + z &= 0 \\ Ru + B^T\lambda - G^T\mu &= 0 \\ Sz + \lambda + \mu &= 0 \\ z - Gu - H\bar{x} &= 0 \ . \end{aligned} \tag{33}$$

In decomposed form :

$$\begin{aligned} A_i\bar{x}_i + B_i u_i + z_i &= 0 \\ R_i u_i + B_i^T\lambda_i - \sum G_{ji}^T\mu_j &= 0 \\ S_i z_i + \lambda_i + \mu_i &= 0 \\ z_i - \sum G_{ij} u_j - \sum H_{ij}\bar{x}_j &= 0 \ . \end{aligned} \tag{34}$$

From (iii), (iv)

$$\begin{aligned} u_i &= -R_i^{-1} B_i^T \lambda_i + R_i^{-1} \sum G_{ji}^T \mu_j \\ \mu_i &= -\lambda_i - S_i \sum G_{ij} u_j - S_i \sum H_{ij} \bar{x}_j \end{aligned} \tag{35}$$

and again it may be seen that μ_j in (35)(i) requires parameters from other subsystems if (35)(ii) is to be solved. Alternatively, working on the aggregate system (33)

$$u = -R^{-1} B^T\lambda + R^{-1} G^T\mu$$

$$z = H\bar{x} - GR^{-1} B^T\lambda + GR^{-1}G^T\mu.$$

Hence(33)(i) and (iii) give

$$A\bar{x} - BR^{-1}B^T\lambda + BR^{-1} G^T\mu + H\bar{x} - GR^{-1}B^T\lambda + GR^{-1}G^T\mu = 0$$

$$SH\bar{x} - SGR^{-1}B^T\lambda + SGR^{-1}G^T\mu + \lambda + \mu = 0$$

or
$$\begin{bmatrix} -(B+G)R^{-1}B^T & (B+G)R^{-1}G^T \\ I-SGR^{-1}B^T & I+SGR^{-1}G^T \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} -(A+H)\bar{x} \\ -SH\bar{x} \end{bmatrix} \qquad (36)$$

The presence of interaction parameters G and G^T on LHS of (36) preclude decomposition.

(iii) λ as coordinating variable

In this case the infimal problem is to solve(21) apart from $\frac{\partial L}{\partial \lambda} = 0$, i.e. (ii)-(v), given $\bar{\lambda}$:

$$\begin{aligned} Qx + A^T\bar{\lambda} - H^T\mu &= b \\ Ru + B^T\bar{\lambda} - G^T\mu &= 0 \\ Sz + \bar{\lambda} + \mu &= 0 \\ z - Gu - Hx &= 0 \ ; \end{aligned} \qquad (37)$$

i.e.
$$z = Gu + Hx$$

$$\mu = -SGu - SHx - \bar{\lambda}$$

so that

$$\begin{aligned} (Q + H^TSH)x + H^TSGu &= -(A+H)^T\bar{\lambda} + b \\ G^TSHx + (R + G^TSG)u &= -(B+G)^T\bar{\lambda} \ . \end{aligned} \qquad (38)$$

Again the presence of H, H^T, G and G^T on the LHS of (38) precludes decomposition.

(iv) u as coordination variable

The infimal problem here is to solve (21) apart from $\frac{\partial L}{\partial u} = 0$, i.e. (i), (ii), (iv), (v), given $\bar{u}$:

$$Ax + B\bar{u} + z = 0$$

$$Qx + A^T\lambda - H^T\mu = b$$

$$Sz + \lambda + \mu = 0$$

$$z - G\bar{u} - Hx = 0 \quad ;$$

i.e.

$$z = G\bar{u} + Hx$$

$$\mu = -Sz - \lambda = -SG\bar{u} - SHx - \lambda \ ,$$

so that

$$(A+H)x + (B+G)\bar{u} = 0$$

$$(Q+H^TSH)x + (A^T+H^T)\lambda + H^TSG\bar{u} = b \qquad (39)$$

and again this is indecomposable.

Hence it is clear that μ must be included as coordinating variable in any successful coordination procedure.

(3) Combination of coordinating variables

Although taking μ as coordinating variable gives a viable decomposition scheme, further combinations are attractive. In particular the two 'interconnection variables' μ and z, one the dual of the other, form a natural pair; and the two multipliers μ and λ, each in the dual space, form another natural pair. Taking z and μ together can be called primal/dual coordination, while taking μ and λ together can be called 'coordination through price instruments', since λ and μ act as prices.

(i) primal/dual

In this scheme the interaction variables $\binom{z}{\mu}$ are set at the supremal level and used as coordinating variables. The infimal problem is then to solve (21) apart from $\frac{\partial L}{\partial z} = \frac{\partial L}{\partial \mu} = 0$, i.e. (i)-(iii), for given $\bar{z}$, $\bar{\mu}$:

$$\left.\begin{aligned} Ax + Bu + \bar{z} &= 0 \\ Qx + A^T\lambda - H^T\bar{\mu} &= b \\ Ru + B^T\lambda - G^T\bar{\mu} &= 0 \end{aligned}\right\} \qquad (40)$$

i.e. $u = -R^{-1}B^T\lambda + R^{-1}G^T\bar{\mu}$.

So

$$Ax - BR^{-1}B^T\lambda = -BR^{-1}G^T\bar{\mu} - \bar{z}$$

$$Qx + A^T\lambda = H^T\bar{\mu} + b \ ,$$

i.e.

$$\begin{pmatrix} A & -BR^{-1}B^T \\ Q & A^T \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -I & -BR^{-1}G^T \\ 0 & H^T \end{pmatrix} \begin{pmatrix} \bar{z} \\ \bar{\mu} \end{pmatrix} + \begin{pmatrix} 0 \\ b \end{pmatrix} \qquad (41)$$

The system (41) is readily decomposed as

$$\begin{pmatrix} A_i & -B_iR_i^{-1}B_i^T \\ Q_i & A_i^T \end{pmatrix} \begin{pmatrix} x_i \\ \lambda_i \end{pmatrix} = \begin{pmatrix} -B_iR_i^{-1}\sum G_{ji}^T\,\bar{\mu}_j - \bar{z}_i \\ \sum H_{ji}^T\,\bar{\mu}_j + b_i \end{pmatrix}, \qquad (42)$$

which can be solved for x_i, λ_i, u_i, $i = 1...N$, if the matrix

$$\begin{pmatrix} A_i & -B_iR_i^{-1}B_i^T \\ Q_i & A_i^T \end{pmatrix}$$

is invertible. This is certainly the case, for compare this matrix with that obtained in scheme (1) above, viz

$$\begin{bmatrix} A_i & -B_i R_i^{-1} B_i^T - S_i^{-1} \\ Q_i & A_i^T \end{bmatrix}$$

which was invertible. Hence the special case S=0 is also invertible. Note too the effect of making z a supremal variable is to supress the matrix S from subsystem inversions.

The gradient of the Lagrangian with respect to $\begin{pmatrix} z \\ \mu \end{pmatrix}$ is

$$\nabla L = \begin{bmatrix} Sz + \lambda + \mu \\ z - Gu - Hx \end{bmatrix}$$

so an updating procedure for $\begin{pmatrix} z \\ \mu \end{pmatrix}$ could be

$$\begin{pmatrix} z \\ \mu \end{pmatrix}^{k+1} = \begin{pmatrix} z \\ \mu \end{pmatrix}^{k} - \varepsilon \begin{bmatrix} Sz^k + \lambda^k + \mu^k \\ z^k - Gu^k - Hx^k \end{bmatrix} . \qquad (43)$$

The information flow for the scheme is as follows:

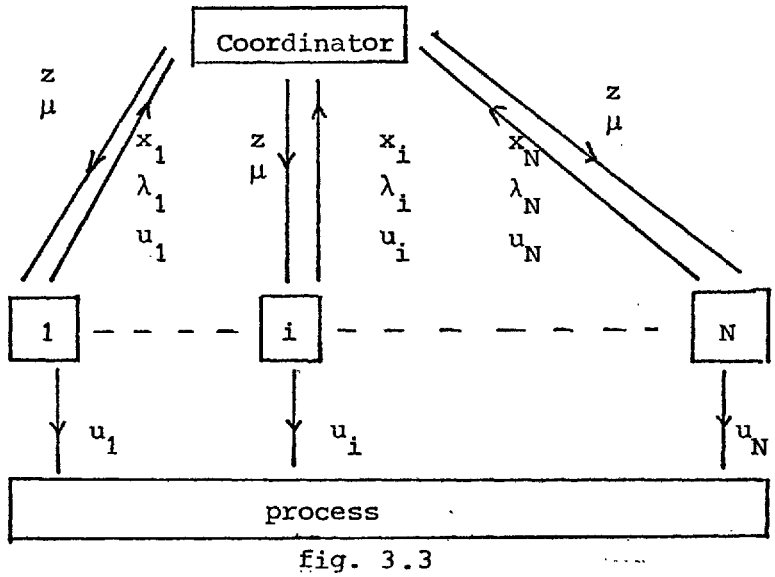


fig. 3.3

A successive approximations update procedure, discussed below, makes this the steady-state analogue of the well-known Takahara scheme.

(ii) Mixed multiplier coordination ["price coordination"]

In this scheme the multipliers $\begin{pmatrix}\lambda\\ \mu\end{pmatrix}$ are set at the supremal level and used as coordinating variables. The infimal problem is then to solve (21) apart from $\frac{\partial L}{\partial \lambda} = \frac{\partial L}{\partial \mu} = 0$, i.e. (ii)-(iv), for given $\bar{\lambda},\bar{\mu}$:

$$\begin{aligned} Qx + A^T\bar{\lambda} - H^T\bar{\mu} &= b \\ Ru + B^T\bar{\lambda} - G^T\bar{\mu} &= 0 \\ Sz + \bar{\lambda} + \bar{\mu} &= 0 \quad , \end{aligned} \tag{48}$$

i.e.
$$\begin{aligned} z &= -S^{-1}(\bar{\lambda}+\bar{\mu}) \\ u &= -R^{-1}(B^T\bar{\lambda}-G^T\bar{\mu}) \\ x &= -Q^{-1}(A^T\bar{\lambda}-H^T\bar{\mu})+Q^{-1}b , \end{aligned}$$

or in decomposed form, for i = i-N ,

$$\begin{aligned} z_i &= -S_i^{-1}(\bar{\lambda}_i+\bar{\mu}_i) \\ u_i &= -R_i^{-1}(B_i^T\bar{\lambda}_i-\Sigma G_{ji}^T\bar{\mu}_j) \\ x_i &= -Q_i^{-1}(A_i^T\bar{\lambda}_i-\Sigma H_{ji}^T\bar{\mu}_j) + Q_i^{-1}b_i . \end{aligned} \tag{49}$$

It is interesting to note that this scheme gives rise to a totally decoupled infimal problem.

The gradient of the Lagrangian with respect to $\begin{pmatrix}\lambda\\ \mu\end{pmatrix}$ is

$$\nabla L = \begin{bmatrix} Ax + Bu + z \\ -Hx - Gu + z \end{bmatrix}_{\lambda},$$

so an updating procedure for $\begin{pmatrix}\lambda\\ \mu\end{pmatrix}$ could be

$$\begin{pmatrix}\lambda\\ \mu\end{pmatrix}^{k+1} = \begin{pmatrix}\lambda\\ \mu\end{pmatrix}^{k} - \varepsilon \begin{bmatrix} Ax^k + Bu^k + z^k \\ -Hx^k - Gu^k + z^k \end{bmatrix} . \tag{50}$$

The information flow for the scheme is as follows:

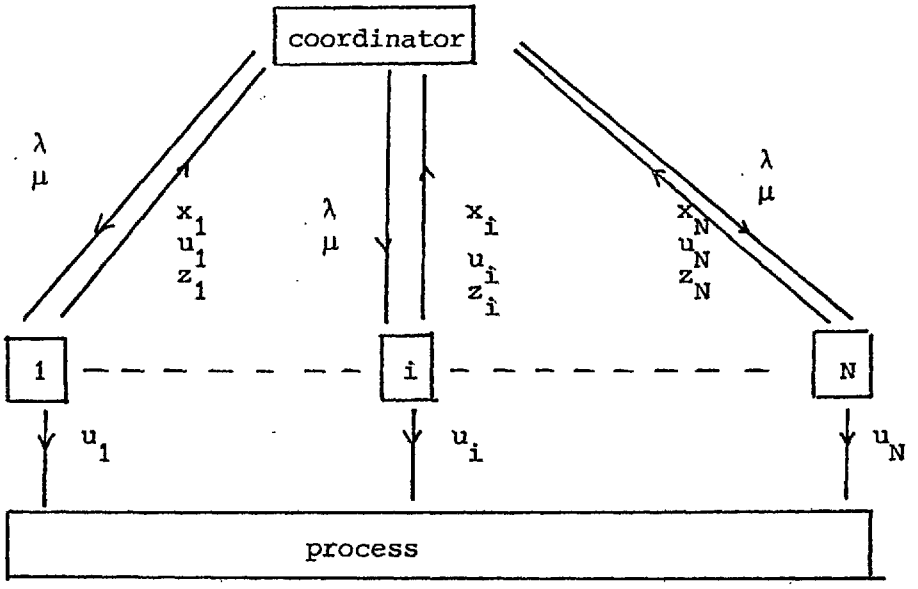


fig. 3.4

Taking combinations of other variables is fruitless from the point of view of producing decomposition schemes. One likely combination might be thought to be $\begin{pmatrix} x \\ z \end{pmatrix}$, since this would set u at the infimal level through

$$\bar{z} = Gu + H\bar{x} \ .$$

But it turns out that this scheme gives an *indeterminate* infimal problem. For consider the infimal equations obtained if $\begin{pmatrix} x \\ z \end{pmatrix}$ is set at the supremal level. They are

$$\begin{aligned} A\bar{x} + Bu + \bar{z} &= 0 \\ Ru + B^T\lambda - G^T\bar{\mu} &= 0 \\ \bar{z} - Gu - H\bar{x} &= 0 \end{aligned} \qquad (51)$$

This system cannot be solved unless $\bar{x}$ and $\bar{z}$ are carefully chosen, so that, for example,

$$A\bar{x} \in \mathcal{R}(B).$$

There is no reason in general why this should be the case.

Similarly taking $\begin{pmatrix} x \\ u \end{pmatrix}$ at supremal level and setting z at the infimal level through

$$z = G\bar{u} + H\bar{x}$$

leaves an indeterminate infimal problem. The equations to be solved in this case are

$$\begin{aligned} A\bar{x} + B\bar{u} + z &= 0 \\ z &= G\bar{u} + H\bar{x} \\ Sz + \lambda + \mu &= 0 \quad , \end{aligned} \tag{52}$$

so that from (52)(i) and (ii)

$$(A+H)\bar{x} + (B+G)\bar{u} = 0 \quad ,$$

which is a restriction on the choice of $\bar{x}$, $\bar{u}$. Further,

$$\lambda + \mu = -SG\bar{u} - SH\bar{u}$$

is indeterminate in λ and μ.

Other combinations are possible, but we shall not pursue them further.

§4. Supremal coordination schemes : decomposition based on successive approximations

In the context of the schemes discussed above, a gradient co - ordination stategy was given based on the steepest descent algorithm. Alternative gradient schemes are of course available, such as conjugate gradient, Fletcher-Reeves, and Newton-Raphson.

An alternative coordination strategy may be based on successive approximations, or "reinjection" as it is called in the multilevel systems literature.

Supremal coordination strategies based on successive approximation have been discussed in decomposition/coordination literature since the publication of Takahara's "reinjection" update scheme for the linear quadratic regulator in 1965. However the discussion has never been satisfactory, and to this day no general convergence proof of the Takahara scheme is available. However formulation of the Lagrangian necessary conditions (21) :

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial u} = \frac{\partial L}{\partial z} = \frac{\partial L}{\partial \lambda} = \frac{\partial L}{\partial \mu} = 0$$

as the linear system

$$Av = b$$

suggests an obvious procedure.

If we apply successive approximations, or Jacobi iteration, to this system we proceed as follows.

Let $A = K + E$

where K is an 'approximation' to A, and E is the 'error'. The matrix K will be chosen judiciously to lead to a decomposition scheme. Then

$$(K+E)v = b$$

$$\Rightarrow \qquad v = -K^{-1}Ev + K^{-1}b$$

which suggests the iteration formula

$$v^{k+1} = -K^{-1}E\, v^{k} + K^{-1}b \qquad (54)$$

or $$Kv^{k+1} = -Ev^{k} + b.$$

As discussed in e.g. Young [58], or Fadeev and Fadeeva [61], a sufficient condition for convergence of this scheme is that

$$\| K^{-1}E \| < 1 ,$$

i.e. that $K^{-1}E$ be a contraction map.

It turns out that K can be chosen so as to yield the well-known Takahara update formulae. *It can also be chosen to lead to decomposition schemes not discussed previously in the literature.* This is one of the principle attractions of adopting this general approach to 'reinjection' supremal coordination schemes.

(1) Takahara scheme

Consider (21a).

$$\text{Set } K = \begin{pmatrix} A & -BR^{-1}B^{T} \\ Q & A^{T} \end{pmatrix} , \qquad E = \begin{pmatrix} H & -BR^{-1}G^{T} - GR^{-1}(B+G)^{T} \\ 0 & H^{T} \end{pmatrix} .$$

.... (55)

Writing this scheme in full, ie substituting for K,E in (54) we have that

$$\begin{pmatrix} A & -BR^{-1}B^T \\ Q & A^T \end{pmatrix} \begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \end{pmatrix} = - \begin{pmatrix} H & -BR^{-1}G^T - GR^{-1}(B+G)^T \\ 0 & H^T \end{pmatrix} \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix} + \begin{pmatrix} 0 \\ b \end{pmatrix} \quad \ldots\ldots (56)$$

which is easily seen to decompose into the N problems

$$A_i x_i^{k+1} - B_i R_i^{-1} B_i^T \lambda_i^{k+1} = - \sum_{j=1}^{N} H_{ij} x_j^k - B_i R_i^{-1} \sum_{j=1}^{N} G_{ji}^T \lambda_j^k$$

$$+ \sum_{j=1}^{N} G_{ij} R_j^{-1} \sum_{\ell=1}^{N} (B_j^T + G_{\ell j}^T) \lambda_\ell^k \qquad (57)\ (i)$$

$$Q_i x_i^{k+1} + A_i^T \lambda_i^{k+1} = - \sum_{j=1}^{N} H_{ji}^T \lambda_j^k + b_i, \qquad i = 1,\ldots N. \qquad (57)\ (ii)$$

We may derive this scheme from the following 2-level approach, using "reinjection", to the solution of the necessary conditions(21).

Supremal level : Update $\begin{pmatrix} z \\ \mu \end{pmatrix}$ by setting

$$\begin{aligned} z^{k+1} &= Gu^k + Hx^k \\ \mu^{k+1} &= -Sz^k - \lambda^k . \end{aligned} \qquad (58)$$

Infimal level : Given $\begin{pmatrix} \bar{z} \\ \bar{\mu} \end{pmatrix}$ solve

$$\begin{aligned} Ax + Bu + \bar{z} &= 0 \\ Qx + A^T\lambda - H^T\bar{\mu} &= b \\ Ru + B^T\lambda - G^T\bar{\mu} &= 0 . \end{aligned} \qquad (59)$$

Setting S = 0, and substituting (58) into (59), we see that this scheme amounts to the system

$$\begin{aligned} Ax^{k+1} + Bu^{k+1} + Gu^k + Hx^k &= 0 \\ Qx^{k+1} + A^T\lambda^{k+1} + H^T\lambda^k &= b \\ Ru^{k+1} + B^T\lambda^{k+1} + G^T\lambda^k &= 0 . \end{aligned} \qquad (60)$$

Eliminating u^{k+1} and u^k gives

$$Ax^{k+1} - BR^{-1}B^T\lambda^{k+1} - BR^{-1}G^T\lambda^k - GR^{-1}B^T\lambda^k - GR^{-1}G^T\lambda^{k-1} + Hx^k = 0,$$

$$Qx^{k+1} + A^T\lambda^{k+1} + H^T\lambda^k = b, \tag{61}$$

which is (56) again, when we set $\lambda^{k-1} = \lambda^k$.

It is important to note that (58) is an alternative supremal update scheme to the gradient scheme employed in the primal-dual approach in §3, when we set

$$\begin{pmatrix} z^{k+1} \\ \mu^{k+1} \end{pmatrix} = \begin{pmatrix} z^k \\ \mu^k \end{pmatrix} - \varepsilon \begin{pmatrix} Sz^k + \lambda^k + \mu^k \\ z^k - Gu^k - Hx^k \end{pmatrix} . \tag{62}$$

(ii) Set $$K = \begin{pmatrix} A & 0 \\ 0 & A^T \end{pmatrix}, \quad E = \begin{pmatrix} H & -(B+G)R^{-1}(B+G)^T \\ Q & H^T \end{pmatrix} \tag{63}$$

This scheme decouples the global necessary conditions into independent equations; thereby simplifying their solutions. Writing the scheme in full, we have

$$\begin{pmatrix} A & 0 \\ 0 & A^T \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix}^{k+1} + \begin{pmatrix} H & -(B+G)R^{-1}(B+G)^T \\ Q & H^T \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix}^k = \begin{pmatrix} 0 \\ b \end{pmatrix}$$

... (64)

which decomposes as

$$A_i x_i^{k+1} + \Sigma H_{ij} x_j^k - B_i R_i^{-1} B_i^T \lambda_i^k - B_i R_i^{-1} \Sigma G_{ji}^T \lambda_j^k$$

$$- \Sigma G_{ij} R_j^{-1} B_j^T \lambda_j^k - \Sigma G_{ij} R_j^{-1} \Sigma G_{\ell j}^T \lambda_\ell^k = 0 \tag{65}$$

$$A_i^T \lambda_i^{k+1} + Q_i x_i^k + \Sigma H_{ji}^T \lambda_j^k = b_i . \tag{66}$$

It may be seen that λ_i^{k+1} does not enter into (65) and x_i^{k+1} does not enter into (66). This scheme offers attractive possibilities

for on-line decentralized regulation ; each subsystem independently updates x(or λ) and these updates are then conveyed to the rest of the system via interconnections. Hence the need for an updating supremal level 'coordinator ' vanishes.

$$\text{(iii)} \quad K = I, \quad E = \begin{bmatrix} A + H - I & -(B+G)R^{-1}(B+G)^T \\ Q & (A+H)^T - I \end{bmatrix}. \tag{67}$$

Writing this scheme in full, we have

$$\begin{aligned} x^{k+1} &= -(A+H-I)x^k + (B+G)R^{-1}(B+G)^T\lambda^k \\ \lambda^{k+1} &= -Qx^k - (A+H)^T\lambda^k + b, \end{aligned} \tag{68}$$

which again may be decomposed as

$$\begin{aligned} x_i^{k+1} &= -A_i x_i^k - x_i^k - \Sigma H_{ij} x_j^k + B_i R_i^{-1} B_i^T \lambda_i^k + B_i R_i^{-1} \Sigma G_{ji}^T \lambda_j^k \\ &\quad + \Sigma G_{ij} R_i^{-1} B_i \lambda_i^k - \Sigma G_{ij} R_i^{-1} \Sigma G_{ij}^T \lambda_j^k = 0 \end{aligned} \tag{69}$$

$$\lambda_i^{k+1} = -Q_i x_i^k - A_i^T \lambda_i^k - \lambda_i^k - \Sigma H_{ji}^T \lambda_j^k + b_i \tag{70}$$

Scheme (iii) might be parameterized to give the possibility of achieving faster convergence by setting :

$$\text{(iv)} \quad K = \gamma I, \quad E = \begin{bmatrix} A+H - \gamma I & -(B+G)R^{-1}(B+G)^T \\ Q & (A+H)^T - \gamma I \end{bmatrix}. \tag{71}$$

Writing this scheme in full, we have

$$\begin{aligned} \gamma x^{k+1} &= -(A+H)x^k - \gamma x^k + (B+G)R^{-1}(B+G)^T\lambda^k \\ \gamma\lambda^{k+1} &= -Qx^k - (A+H)^T\lambda^k - \gamma\lambda^k + b, \end{aligned} \tag{72}$$

which may be decomposed as

$$\begin{aligned} \gamma x_i^{k+1} &= -A_i x_i^k - \gamma x_i^k - \Sigma H_{ij} x_j^k + B_i R_i^{-1} B_i^T \lambda_i^k + B_i R_i^{-1} \Sigma G_{ji}^T \lambda_j^k \\ &\quad + \Sigma G_{ij} R_j^{-1} B_j^T \lambda_j^k + \Sigma G_{ij} R_j^{-1} \Sigma G_{lj}^T \lambda_l \\ \gamma\lambda_i^{k+1} &= -Q_i x_i^k - A_i^T \lambda_i^k - \gamma\lambda_i^k - \Sigma H_{ji}^T \lambda_j^k + b_i. \end{aligned} \tag{73}$$

We now consider the convergence of the successive approximations schemes. Returning to (54), we have the general form

$$v^{k+1} = -K^{-1}E\, v^{k} + K^{-1} b . \tag{74}$$

We noted that a sufficient condition of convergence is

$$\| K^{-1}E \| < 1 \tag{75}$$

i.e. that $K^{-1}E$ be a contraction operator.

Let us translate (75) into specific cases. For the most important cases, that of the static "Takahara" scheme and the decoupled scheme, we have :

(1) Takahara scheme :

$$K = \begin{pmatrix} A & -BR^{-1}B^{T} \\ Q & A^{T} \end{pmatrix} , \quad E = \begin{pmatrix} H & -BR^{-1}G^{T} - GR^{-1}(B+G)^{T} \\ 0 & H^{T} \end{pmatrix} .$$

..... (76)

$$\text{Hence } K^{-1}E = \begin{pmatrix} A & -BR^{-1}B^{T} \\ Q & A^{T} \end{pmatrix}^{-1} \begin{pmatrix} H & -BR^{-1}G^{T} - GR^{-1}(B+G)^{T} \\ 0 & H^{T} \end{pmatrix}$$

..... (77)

It may be seen that (77) is too complicated for checking on convergence, and is of theoretical importance only. It does not seem to admit of any simple interpretation in terms of system parameters.

(2) Decoupled scheme

$$K = \begin{pmatrix} A & 0 \\ 0 & A^{T} \end{pmatrix} , \quad E = \begin{pmatrix} H & -(B+G)R^{-1}(B+G)^{T} \\ Q & H^{T} \end{pmatrix} \tag{78}$$

$$\text{Hence } K^{-1}E = \begin{pmatrix} A^{-1} & 0 \\ 0 & A^{T-1} \end{pmatrix} \begin{pmatrix} H & -(B+G)R^{-1}(B+G)^{T} \\ Q & H^{T} \end{pmatrix}$$

$$= \begin{pmatrix} A^{-1}H & -A^{-1}(B+G)R^{-1}(B+G)^{T} \\ A^{T-1}Q & A^{T-1}H^{T} \end{pmatrix} \tag{79}$$

The condition $\|K^{-1}E\| < 1$ here translates into more productive conditions.

Since in general it is true that

$$[\|A\| < \frac{1}{\lambda},\ \|B\| < \frac{1}{\lambda},\ \|C\| < \frac{1}{\lambda},\ \|D\| < \frac{1}{\lambda}] \Rightarrow \left\| \begin{matrix} A & B \\ C & D \end{matrix} \right\| < 1 \qquad (80)$$

we have that a set of sufficient conditions for convergence is

$$\|A^{-1}H\| < \frac{1}{\lambda} \qquad (81)$$

$$\|A^{-1}(B+G)R^{-1}(B+G)^T\| < \frac{1}{\lambda} \qquad (82)$$

$$\|A^{T^{-1}}Q\| < \frac{1}{\lambda} \qquad (83)$$

$$\|A^{T^{-1}}H^T\| < \frac{1}{\lambda} \qquad (84)$$

Note that (81) and (84) are a form of "diagonal dominance" since A represents diagonal elements and H the off-diagonal elements.

Finally let us consider acceleration of the successive approximations schemes by recourse to Gauss-Seidel iteration. This scheme uses updates of the iterated variable as soon as they become available.

Write

$$Av = b \qquad (85)$$

as

$$(K + L + U)v = b \qquad (86)$$

where U is upper block triangular

and L is lower block triangular

and $E = L+U$.

Then $(K+L)v = -Uv+b$

so we have the iteration formula

$$v^{k+1} = -(K+L)^{-1}Uv^k + (K+L)^{-1}b \qquad (88)$$

where we must ensure that (K+L) is in fact invertible.

This is certainly the case when K is block-diagonal with invertible blocks.

We may see that the general block-diagonal system

$$\begin{pmatrix} K_1 & & & & 0 \\ L_{21} & K_2 & & & \\ L_{31} & L_{32} & K_3 & & \\ & & \ddots & \ddots & \\ L_{N_1} & L_{N_2} & \cdots & \cdots & K_N \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \vdots \\ \omega_N \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_N \end{pmatrix}$$

..... (89)

can be solved sequentially :

$$\begin{aligned} \omega_1 &= -K_1^{-1}(-\gamma_1) \\ \omega_2 &= -K_2^{-1}(L_{21}\omega_1 - \gamma_2) \\ &\vdots \\ \omega_N &= -K_N^{-1}(L_{N_1}\omega_1 + L_{N_2}\omega_2 + \ldots + L_{N_{N-1}}\omega_{N-1} - \gamma_N) \end{aligned}$$

provided K_i, i = 1...N,are themselves invertible.

Now for the totally decentralized case (ii) above, where

$$K = \begin{pmatrix} A & 0 \\ 0 & A^T \end{pmatrix},$$

these conditions are certainly met, i.e. A and A^T are assumed invertible.

In this case we have that

$$L = \begin{pmatrix} 0 & 0 \\ Q & 0 \end{pmatrix}$$

$$U = \begin{pmatrix} H & -(B+G)R^{-1}(B+G)^T \\ 0 & H^T \end{pmatrix}$$

so the full iteration formula is

$$\begin{pmatrix} A & 0 \\ Q & A^T \end{pmatrix}\begin{pmatrix} x \\ \lambda \end{pmatrix}^{k+1} = \begin{pmatrix} H & -(B+G)R^{-1}(B+G)^T \\ 0 & H^T \end{pmatrix}\begin{pmatrix} x \\ \lambda \end{pmatrix}^{k} + \begin{pmatrix} 0 \\ b \end{pmatrix}$$

..... (90)

which again decomposes as

$$A_i x_i^{k+1} = \sum_{j=1}^{N} H_{ij} x_j^k - B_i R_i^{-1}\left(B_i^T \lambda_i^k + \sum G_{ji}^T \lambda_j^k\right) - \sum G_{ij} R_j^{-1}\left(B_j^T \lambda_j + \sum G_{ej}^T \lambda_e^k\right)$$

$$Q_i x_i^{k+1} + A_i^T \lambda_i^k = \sum H_{ji}^T \lambda_j^k + b_i \quad .$$

By contrast it is interesting to observe that the Takahara Scheme *is already in Gauss-Seidel form*, for in this case we have that

$$K = \begin{pmatrix} A & -BR^{-1}B^T \\ Q & A^T \end{pmatrix}$$

$$E = \begin{pmatrix} H & -BR^{-1}G^T - GR^{-1}(B+G)^T \\ 0 & H^T \end{pmatrix},$$

so that $L = 0$,

and $U = E$.

Putting $L = 0$ in (88) gives

$$v^{k+1} = -K^{-1}Uv^k + K^{-1}b$$

which is the Jacobi iteration formula again with $U = E$.

It should be noted that although K is not diagonal, it is decomposable, and hence easy to invert at subsystem level. Hence the spirit of Jacobi iteration is preserved.

It is an interesting comment on the 'efficiency' of the Takahara Scheme, that it is already in accelerated Gauss-Seidel format. This observation seems never to have appeared in the literature before.

We could continue and introduce relaxation parameters, but this would take us too far afield.

§5 Gradient methods and interaction prediction

(1) Introduction and motivation

Hitherto we have focussed our attention on the Lagrange necessary conditions of optimality, and allowed multipliers (λ,μ) to play a key role in decomposition and coordination schemes. We now adopt a different perspective altogether, and make the system gradient the focus of our attention. We shall assign the task of computing the system gradient to a supremal level; we shall then look for gradient-based algorithms that generate decomposition schemes, allowing the algorithm computations to be carried out in parallel by subsystems each supplied with partial information on the gradient.

This is a fruitful line of attack, that has been initiated only recently by Cohen [6,7] . Not only does it yield new decomposition schemes in its own right, but it throws new light on the Lagrangian schemes. These can be reinterpreted as computing in a decentralized fashion. the 'local' approximations to system state and costate variables. Hence they carry a lighter computational load than gradient-based schemes - but at a price, which we discuss below.

Recall that for the general quadratic cost

$$J(u) = a + \beta^T u + \tfrac{1}{2}u^T C u \tag{91}$$

we have the gradient

$$g(u) = \beta + Cu \tag{92}$$

and optimal control

$$u^* = -C^{-1}\beta \ , \tag{93}$$

given by setting $g(u) = 0$

In general, the inversion of C may be a difficult task, and a number of algorithms have been devised which replace C^{-1} in (93) by

K^{-1}, where K is easy to invert. In particular, if K is block diagonal it will be easy to invert, *and* give rise to a decomposition scheme.

Three such algorithms are well-known : the Newton-Raphson (K^{-1}), the Contraction Mapping (K^{-1}), and Conjugate Gradient (K^{-1}) algorithms.

The CM(K^{-1}) scheme is easily derived. Set

$$C = K + E \tag{94}$$

and write the optimality criterion

$$g(u^*) = 0$$

as

$$\beta + (K+E)u^* = 0$$

i.e.

$$u^* = -K^{-1}E\, u^* - K^{-1}\beta . \tag{95}$$

This suggests the iterative scheme

$$u^{k+1} = -K^{-1}E\, u^k - K^{-1}\beta, \tag{96}$$

which converges if $\| K^{-1}E\| < 1$. (97)

The Newton-Raphson scheme for minimizing (91) is

$$u^{k+1} = u^k - \left(\frac{d^2J}{du^2}\,(u^k) \right)^{-1} \frac{dJ}{du}(u^k) \tag{98}$$

or

$$u^{k+1} = u^k - C^{-1}\, g(u^k).$$

Approximating C^{-1} by K^{-1} we have the NR(K^{-1}) scheme :

$$u^{k+1} = u^k - K^{-1}g(u^k). \tag{99}$$

It is easily seen that in the quadratic case, the CM(K^{-1}) and NR(K^{-1}) algorithms are identical (for the same K). Indeed, we have that

$$\begin{aligned} u^{k+1} &= u^k - K^{-1}g(u^k) && \text{NR}(K^{-1}) \\ &= u^k - K^{-1}(\beta+Cu^k) && \text{by (92)} \\ &= u^k - K^{-1}(\beta+(K+E)u^k) && \text{by (94)} \\ &= u^k - K^{-1}\beta - u^k - K^{-1}Eu^k \end{aligned}$$

which is (96) again. Hence (97) is also a convergence condition for the $NR(K^{-1})$ algorithm. We shall derive equivalent forms below.

Finally the $CG(K^{-1})$ algorithm, first given by Hestenes[46], is as follows:

Step 0. $s^0 = -K^{-1}g^0, \quad k = 0$

1. $u^{k+1} = u^k + \omega s^k$, with ω as the optimal step length. (95)

2. $s^{k+1} = -K^{-1}g^{k+1} + \beta s^k$

$$\beta = \frac{\langle g^{k+1}, K^{-1}g^{k+1}\rangle}{\langle g^k, K^{-1}g^k\rangle}$$

3. Set k = k+1, and go to 1.

Clearly, if K = I then the CG(I) algorithm is just the ordinary CG algorithm. The $CG(K^{-1})$ scheme is attractive because it converges for all $K = K^T > 0$, and moreover it tends to converge faster than the $NR(K^{-1})/CM(K^{-1})$ schemes on quadratic costs.

It is remarkable that if we start from a different viewpoint, and employ Mesarovic's interaction operator to modify local subsystem criteria [each modification ensuring that the locally computed sub-gradients, when aggregated, give the true system gradient] then we arrive at a scheme identical with the $NR(K^{-1})/CM(K^{-1})$, provided

(i) we choose K in a natural way to be specified below;

(ii) we coordinate the IP algorithm through 'reinjection'.

This identity was first established by Cohen and Bernhard [7] using heuristic arguments based on quadratic expansions. In this section we exploit the transparency of the static LQ regulator to show the identity directly, by looking at the control sequence generated in each case.

(2) Derivation of system gradient

Recall from Chapter 2 that our basic model is

$$\min_{u} G(x,u)$$
$$\text{s.t.} \quad x = P(u) \tag{96}$$

or $x = \bar{P}(u,z)$,

where $z = K(u) = H(u,x)$,

while the static LQ regulator is formulated as

$$\min \tfrac{1}{2}x^T Qx + \tfrac{1}{2}u^T Ru - b^T x$$
$$\text{s.t.} \quad Ax + Bu + z = 0 \tag{97}$$
$$z = Gu + Hx$$

Reforming (97) as (96), we have that

$$x = -A^{-1}Bu - A^{-1}z = \bar{P}(u,z) \tag{98}$$

$$\text{or} \quad x = \hat{W}u - A^{-1}z, \tag{99}$$

while

$$z = Gu + Hx = H(u,x). \tag{100}$$

Overall

$$(A+H)x + (B+G)u = 0 \tag{101}$$

so that

$$x = -(A+H)^{-1}(B+G)u = P(u) \tag{102}$$

$$\text{or} \quad x = Wu, \tag{103}$$

$$\text{and} \quad z = [G - H(A+H)^{-1}(B+G)]u = K(u). \tag{104}$$

Criteria are

$$G(u,x) = \tfrac{1}{2}x^T Qx + \tfrac{1}{2}u^T Ru - b^T x \tag{105}$$

so that

$$J(u) \triangleq G(u,P(u)) = \tfrac{1}{2}u^T[R+W^T QW]u - b^T Wu \tag{106}$$

and the system gradient is

$$g(u) = \frac{dJ}{du} = (R+W^T QW)u - W^T b \tag{107}$$

and $$\frac{d^2J}{du^2} = R + W^T Q W. \tag{108}$$

Alternatively we may derive a 'decoupled' or decentralized version of the cost, which will give a 'decoupled' form of the gradient that will be referred to in the sequel.

Set

$$\bar{J}(u,z) = G(u,\bar{P}(u,z)) \tag{109}$$

$$= \tfrac{1}{2}(u^T\hat{W}^T - z^T A^{-1^T})Q(\hat{W}u - A^{-1}z) + \tfrac{1}{2}u^T Ru - b^T(\hat{W}u - A^{-1}z)$$

$$= \tfrac{1}{2}u^T[R + \hat{W}^T Q\hat{W}]u - \tfrac{1}{2}u^T\hat{W}^T Q A^{-1}z - \tfrac{1}{2}z^T A^{-1^T} Q\hat{W}u$$

$$+ \tfrac{1}{2}z^T A^{-1^T} Q A^{-1}z - b^T\hat{W}u + b^T A^{-1}z \ , \tag{110}$$

so that

$$\hat{g}(u,z) = \frac{\partial\bar{J}}{\partial u} = (R + \hat{W}^T Q\hat{W})u - \hat{W}^T Q A^{-1}z - \hat{W}^T b \tag{111}$$

....

and

$$\frac{\partial^2\bar{J}}{\partial u^2} = R + \hat{W}^T Q\hat{W}. \tag{112}$$

(3) Relations with Lagrangian approach

Alternative formulae for system gradient can be found via the system costate. In the static case this is a rather trivial formulation, but it is of fundamental importance in the dynamic domain. Consider now the overall problem

$$\min \quad G(x,u)$$

$$\text{s.t.} \qquad (A+H)x + (B+G)u = 0 \ .$$

We have the Lagrangian

$$L(x,u,\lambda) = G(x,u) + \lambda^T[(A+H)x + (B+G)u] \tag{113}$$

so that

$$\frac{\partial L}{\partial u} = Ru + (B+G)^T\lambda \ . \tag{114}$$

Now rewrite the system gradient (107) as

$$g(u) = Ru + W^T QWu - W^T b$$

$$= Ru + W^T Qx - W^T b \qquad \text{by (103)}$$

$$= Ru - (B+G)^T (A+H)^{-1^T} Qx + (B+G)^T (A+H)^{-1^T} b \quad \text{by (102)}$$

so that

$$g(u) = Ru + (B+G)^T \lambda \tag{115}$$

if we define

$$\lambda = -(A+H)^{-1^T} Qx + (A+H)^{-1^T} b. \tag{116}$$

Equation (116) is the static analogue of the costate or adjoint equation for dynamic systems. It provides the link between the system gradient (107) and the Lagrangian derivative (114).

Similarly we have the 'decoupled' case,

$$\min \quad G(x,u) \tag{117}$$

$$\text{s.t. } Ax + Bu + z = 0,$$

when z is treated as a 'disturbance' coming from outside each subsystem, i.e.

$$A_i x_i + B_i u_i + z_i = 0 , \qquad i = 1 \ldots N.$$

In this case we have the decoupled Lagrangian

$$\bar{L}(x,u,z,\hat{\lambda}) = G(x,u) + \hat{\lambda}^T (Ax + Bu + z) \tag{118}$$

so that

$$\frac{\partial \bar{L}}{\partial u} = Ru + B^T \hat{\lambda} \quad . \tag{119}$$

Rewrite the 'decoupled' system gradient $\hat{g}(u,z)$ (111) as

$$\hat{g}(u,z) = Ru + \hat{W}^T Q \hat{W} u - \hat{W}^T Q A^{-1} z - \hat{W}^T b$$

$$= Ru + \hat{W}^T Q \hat{x} - \hat{W}^T b \qquad \text{by (99)}$$

$$= Ru - B^T A^{-1^T} Q \hat{x} + B^T A^{-1^T} b \qquad \text{by (98)}$$

so that

$$\hat{g}(u) = Ru + B^T \hat{\lambda} \tag{120}$$

if we define

$$\hat{\lambda} = -A^{-1^T} Q\hat{x} + A^{-1^T} b \tag{121}$$

Again the 'costate' equation (121) provides the link between the decoupled system gradient (111) and the derivative of the decoupled Lagrangian (119). Note that the decoupled formulation lends itself to decomposition :

$$\left.\begin{aligned} \hat{\lambda}_i &= -A_i^{-1^T} Q_i \hat{x}_i + A_i^{-1^T} b_i \\ \hat{x}_i &= -A_i^{-1} B_i u_i - A_i^{-1} z_i \\ \hat{g}_i(u_i) &= R_i u_i + B_i^T \hat{\lambda}_i \end{aligned}\right\} \tag{122}$$

4. Gradient algorithms

We now apply the three gradient algorithms mentioned above to the quadratic cost (106)

$$J(u) = \tfrac{1}{2} u^T [R + W^T Q W] u - b^T W u \tag{123}$$

so that in the general quadratic expression (91) we have

$$\left.\begin{aligned} a &= 0 \\ \beta &= -W^T b \\ C &= R + W^T Q W. \end{aligned}\right\} \tag{124}$$

We propose to approximate C by the decomposable form

$$K = R + \hat{W}^T Q \hat{W}. \tag{125}$$

This suggestion is not new; it was made independently by Allwright [47] in the context of the $CG(K^{-1})$ algorithm, and by Cohen [6,7] in the context of the $NR(K^{-1})/CM(K^{-1})$ algorithm. These suggestions are however placed in a unified framework here for the first time.

(i) $CG(K^{-1})$ algorithm

The update step is

$$u^{k+1} = u^k + \omega s^{k+1} \tag{126}$$

where $$s^{k+1} = -K^{-1}g^{k+1} + \frac{\langle g^{k+1}, K^{-1}g^{k+1}\rangle}{\langle g^k, K^{-1}g^k\rangle}\, s^k$$

and $$K = R + \hat{W}^T Q \hat{W} = R + B^T A^{-1^T} Q\, A^{-1} B.$$

Clearly this step can be distributed to the N update steps

$$s_i^{k+1} = -(R_i + B_i^T A_i^{-1^T} Q_i A_i^{-1} B_i)^{-1} g_i^{k+1} + \frac{\langle g_i^{k+1}, (R_i + B_i^T A_i^{-1^T} Q_i A_i^{-1} B_i)^{-1} g_i^{k+1}\rangle}{\langle g_i^k, (R_i + B_i^T A_i^{-1^T} Q_i A_i^{-1} B_i)^{-1} g_i^k\rangle}\, s_i^k, \tag{127}$$

$$i = 1, \ldots N,$$

where g_i is supplied to each subsystem by the coordinator. At the supremal level the system gradient is computed either directly in terms of u, by (107) :

$$g(u) = [R + (B+G)^T (A+H)^{-1} Q (A+H)^{-1} (B+G)]u + (B+G)^T (A+H)^{-1^T} b$$

or indirectly in terms of x, λ, u, by (115), (116) and (102)

$$\left.\begin{aligned} g(u) &= Ru + (B+G)^T \lambda \\ \lambda &= -(A+H)^{-1^T} Qx + (A+H)^{-1^T} b \\ x &= -(A+H)^{-1} (B+G) u. \end{aligned}\right\} \tag{128}$$

Again it should be emphasised that although the indirect computation of system gradient, using x and λ, seems an unnecessary complication here, in the dynmaic case it is fundamental.

Hence we have the following exchange of information between supremal and infimal levels :

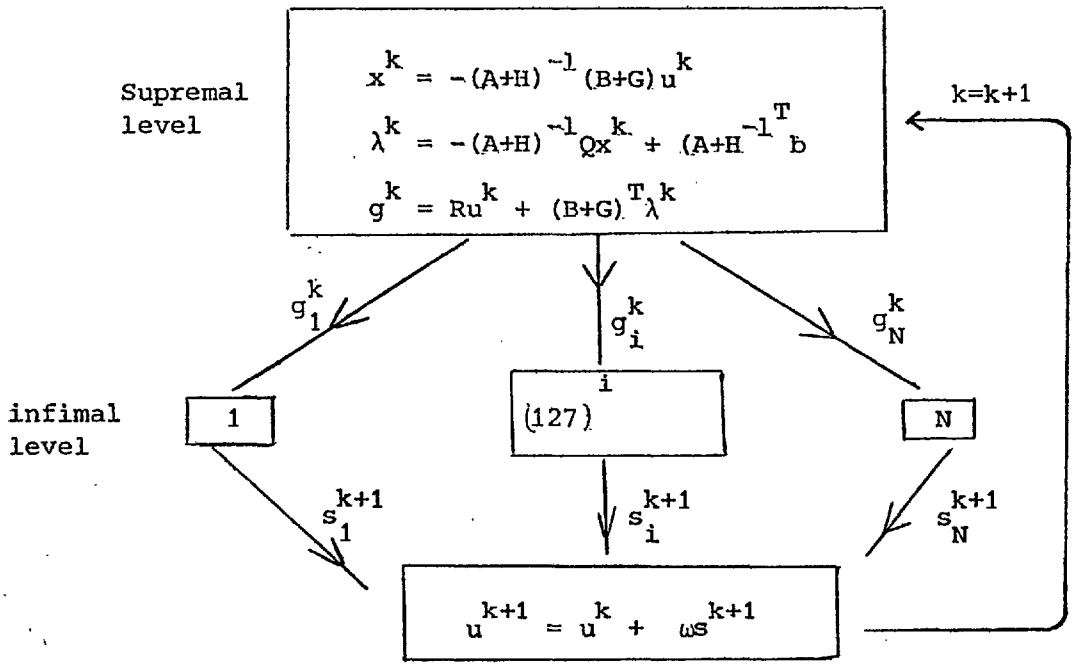


fig. 5

In (47) Allwright introduced a simple scheme for evaluating $K^{-1}g(u^k)$ in decoupled fashion, but in the static case inversion of the matrix K presents no problems, so we shall postpone discussion of this to the dynamic case.

(ii) NR(K^{-1}/CM(K^{-1}) algorithm.

Using the NR(K^{-1}) formulation, we have the update step

$$u^{k+1} = u^k - K^{-1}g(u^k)$$

or
$$u^{k+1} = u^k - (R+\hat{W}^T Q\hat{W})^{-1}\, g(u^k) \qquad (129)$$

which again may be decomposed into the N updates :

$$u_i^{k+1} = u_i^k - (R_i + B_i^T A_i^{-1}{}^T Q_i A_i^{-1} B_i)^{-1}\, g_i^k \qquad (130)$$

where g_i^k is supplied by the coordinator.

The system gradient may be computed either directly or indirectly, and the exchange of information between supremal and infimal level will be exactly as in fig 5 with the exception that update expressions 127(i),(ii) are replaced by (130), i = 1 ...N_1 and $s = K^{-1}g$.

We now discuss convergence of this scheme. Recall from (97) that a sufficient condition for convergence is that $K^{-1}E$ be a contraction map, i.e.

$$\| K^{-1}E\| < 1. \tag{131}$$

An equivalent condition is that

$$\| K^{-1}C - I\| < 1, \tag{132}$$

for

$$\begin{aligned} K^{-1}C - I &= K^{-1} C - K^{-1}K \\ &= K^{-1}(C - K) \\ &= K^{-1}E. \end{aligned}$$

A closely related condition is

$$\| K^{-\frac{1}{2}}C K^{-\frac{1}{2}} - I\| < 1 \tag{133}$$

which can be shown to be equivalent to

$$K - \tfrac{1}{2}C > 0 \tag{134}$$

Indeed, we have that

$$\begin{aligned} K - \tfrac{1}{2}C > 0 &\iff K^{-\frac{1}{2}}(K-\tfrac{1}{2}C)K^{-\frac{1}{2}} > 0, \text{ since } K^{-\frac{1}{2}} \text{ has full rank} \\ &\iff I - \tfrac{1}{2}K^{-\frac{1}{2}}CK^{-\frac{1}{2}} > 0 \\ &\iff 2I - K^{-\frac{1}{2}}CK^{-\frac{1}{2}} > 0 \\ &\iff I > K^{-\frac{1}{2}}CK^{-\frac{1}{2}} - I \\ &\iff 1 > \| K^{-\frac{1}{2}}CK^{-\frac{1}{2}} - I\|. \end{aligned}$$

Cohen [6,7] claims that (134) is exactly equivalent to (132), but this does not appear to be so. It is interesting to note that the condition (134) can be derived from a completely different argument[still within the context of the general quadratic]. It is shown in Cohen and Bernhard [7] that $J(u^k)$ strictly decreases with k according to

$$J(u^k) - J(u^{k+1}) = \langle u^k - u^*, Z(u^k - u^*)\rangle \tag{135}$$

where $Z = CK^{-1}(K-\frac{1}{2}C)K^{-1}C$, (136)

Consequently the cost decreases iff $Z > 0$, i.e. $K-\frac{1}{2}C > 0$.

We now indicate how this result is obtained. We have that

$$J(u) = a + \langle b,u\rangle + \langle u,Cu\rangle$$

Hence $\frac{dJ}{du}(u^k) = Cu^k + b$

and $\frac{dJ}{du}(u^*) = Cu^* + b = 0$

so that $\frac{dJ}{du}(u^k) = C(u^k-u^*)$, (137)

Then the NR(K^{-1}) update formula is

$$u^{k+1} - u^k = -K^{-1}\frac{dJ}{du}(u^k) = -K^{-1}C(u^k-u^*) \quad (138)$$

Expanding $J(u^{k+1})$ to 2^{ND} order about u^k, we have

$$J(u^{k+1}) = J(u^k) + \langle \nabla J(u^k), u^{k+1}-u^k\rangle + \tfrac{1}{2}\langle u^{k+1}-u^k, \frac{d^2J}{du^2}(u^k)(u^{k+1}-u^k)\rangle$$

and substituting (137) and (138) gives

$$\begin{aligned} J(u^{k+1}) - J(u^k) &= \langle C(u^k-u^*), -K^{-1}C(u^k-u^*)\rangle \\ &\quad +\tfrac{1}{2}\langle K^{-1}C(u^k-u^*), CK^{-1}C(u^k-u^*)\rangle \\ &= \langle u^k-u^*, (\tfrac{1}{2}CK^{-1}CK^{-1}C - CK^{-1}C)(u^k-u^*)\rangle \\ &= \langle u^k-u^*, Z(u^k-u^*)\rangle \end{aligned}$$

with Z given by (136), and so we confirm that

$J(u^{k+1}) - J(u^k) < 0$ provided $C > 0$, $K > 0$ and $K-\frac{1}{2}C > 0$.

We ensure $K > 0$ and $C > 0$ by original convexity assumptions : $Q \geq 0$, $R > 0$.

We now proceed to derive an interaction prediction scheme using interaction operator modification of local criteria. We shall show, using the static analogue of Cohen's argument in the dynamic case [6,7], that this algorithm can be made

identical with the NR$(R + \hat{W}^T Q \hat{W})^{-1}$ scheme given here, provided it is combined with a 'reinjection' coordination step.

(5) Interaction prediction

We now apply the interaction prediction method to the static LQ regulator, using interaction operators to modify local criteria. Recall from chapter 2 §3 that, given a current control u^k, we determine current state $x^k = P(u^k)$ and costate λ^k and current interaction prediction $z^k = H(x^k, u^k)$. The local optimization problems are then solved with this interaction term and criterion modification

$$\begin{aligned} &\min_u \; G(x,u) + \Gamma(u^k)u \\ &\text{s.t.} \quad x = \bar{P}(u,z) \\ &\qquad\; z = K(u^k) \end{aligned} \tag{139}$$

where $\Gamma(u^k) = \frac{\partial \bar{J}}{\partial z}(u^k, K(u^k))^T \frac{dz}{du}(u^k)$. (140)

We specify a simple 'reinjection' coordination algorithm, whereby we take u^{k+1} as $\hat{u}(\alpha^k)$ the optimal control found at step k, for a given interaction prediction α^k.

We proceed to the derivation of the row vector $\Gamma(u)$. We have from the expression (110) for $\bar{J}(u,z)$ that

$$\frac{\partial \bar{J}}{\partial z} = A^{-1^T} Q A^{-1} z - A^{-1^T} Q \hat{W} u + A^{-1^T} b. \tag{141}$$

It will also be useful to derive this indirectly, in the form

$$\frac{\partial \bar{J}}{\partial z}(u,z) = \frac{\partial \bar{P}}{\partial z}(u,z)^T \frac{\partial G}{\partial x}(u, \bar{P}(u,z)). \tag{142}$$

We have that

$$\frac{\partial G}{\partial x}(u,x) = Qx - b \tag{143}$$

and

$$\frac{\partial \bar{P}}{\partial z}(u,z) = -A^{-1} \tag{144}$$

so that substituting (98)

$$x = -A^{-1}Bu - A^{-1}z$$

gives

$$\frac{\partial \bar{P}}{\partial z}(u,z)^T \frac{\partial G}{\partial x}(u,\bar{P}(u,z)) = A^{-1^T} Q[A^{-1}Bu+A^{-1}z] + A^{-1^T}b.$$

which is (141) again.

Next we eliminate z from (141), using (104) :

$$\frac{\partial \bar{J}}{\partial z}(u,.) = A^{-T}QA^{-1}Bu + A^{-1^T}QA^{-1}[G - H(A+H)^{-1}(B+G)]u + A^{-1^T}b. \quad \ldots\ldots \quad (145)$$

Finally from (104) again we have that

$$\frac{dK}{du} = G - H(A+H)^{-1}(B+G) \quad (146)$$

Hence, we evaluate $\Gamma(u)$ from (145) and (146), according to (140) :

$$\begin{aligned}\Gamma(u) &= \frac{\partial \bar{J}}{\partial z}(u)^T \frac{dK}{du} \\ &= \{u^T[B^TA^{-1^T}QA^{-1} + \{G^T - (B+G)^T(A+H)^{-1^T}H^T\}A^{-1^T}QA^{-1}] \\ &\qquad + b^TA^{-1}\}[G-H(A+H)^{-1}(B+G)]\end{aligned} \quad \ldots\ldots \quad (147)$$

Expression (147) is exceedingly clumsy, so we proceed analogously to Cohen [6,7] and obtain a simpler expression for $\Gamma(u)$ in terms of the system state and costate variables x and λ.

We have the following convenient

Lemma (Cohen)

$$\Gamma(\bar{u})u = \langle \bar{\lambda},\ Hx + HA^{-1}z + Gu \rangle \quad (148)$$

where $\bar{u}$ is given control,

$$\bar{x} = -(A+H)^{-1}(B+G)\bar{u}, \quad (149)$$

$$\bar{\lambda} = -(A+H)^{-1^T}Q\bar{x} + (A+H)^{-1^T}b \quad (150)$$

$$x = -A^{-1}Bu \quad (151)$$

Proof

$$\text{Since } \Gamma(u) = \frac{\partial G}{\partial x}^T \frac{\partial \bar{P}}{\partial z} \frac{dK}{du} \tag{152}$$

we have, using (143), (144) and (146) that

$$\Gamma(u) = -(Qx-b)^T A^{-1}[G-H(A+H)^{-1}(B+G)] \tag{153}$$

or

$$\Gamma(u)u = \langle Qx-b, -A^{-1}[G-H(A+H)^{-1}(B+G)]u\rangle \tag{154}$$

which is the form in which we need the modification in (139).

We now appeal to tbe identities

$$-(A+H)^{-1}+A^{-1} = A^{-1}H(A+H)^{-1} = (A+H)^{-1}HA^{-1}, \tag{155}$$

to show that

$$\Gamma(u)u = \langle Qx-b, -(A+H)^{-1}(G-HA^{-1}B)u\rangle, \tag{156}$$

for $-A^{-1}G + A^{-1}H(A+H)^{-1}(B+G)$.

$= -A^{-1}G - [(A+H)^{-1} - A^{-1}](B+G)$ by 155 (i)

$= -A^{-1}G - (A+H)^{-1}B - (A+H)^{-1}G + A^{-1}B + A^{-1}G$

$= [A^{-1} - (A+H)^{-1}]B - (A+H)^{-1}G$

$= (A+H)^{-1}HA^{-1}B - (A+H)^{-1}G$ by 155 (ii)

Hence, rearranging (156)

$$\Gamma(\bar{u})u = \langle -(A+H)^{-1}{}^T(Q\bar{x}-b),\ Gu - HA^{-1}Bu\rangle$$

$$= \langle \bar{\lambda},\ Gu + Hx + HA^{-1}z\rangle$$

where $\bar{\lambda} = -(A+H)^{-1}{}^T(Q\bar{x}-b)$

and $x = -A^{-1}Bu$. ‖

Note the clear distinction between the *system variables* $\bar{u}$, $\bar{x}$, $\bar{\lambda}$ and the *local variables* x,u.

We now write the modified problem (139) in detail as

$$\left.\begin{aligned} &\min_u \tfrac{1}{2}x^TQx + \tfrac{1}{2}u^TRu - b^Tx + \lambda^{k^T}(Hx + Gu) \\ &\text{s.t.} \quad Ax + Bu + z = 0 \\ &\qquad z = Gu^k + Hx^k, \end{aligned}\right\} \tag{157}$$

where we ignore the term $HA^{-1}z$ in the cost since it does not affect minimization with respect to u only.

This decomposes into N interconnected problems

$$\min_{u_i} \tfrac{1}{2}x_i^T Q_i x_i + \tfrac{1}{2}u_i^T R_i u_i - b_i^T x_i + \lambda_i^{k\,T}\Big(\sum_{j\neq i} H_{ij}x_j + \sum_{j\neq i} G_{ij}u_j\Big)$$

$$\text{s.t.} \quad A_i x_i + B_i u_i + z_i = 0 \qquad (158)$$

$$z_i = \sum_{j\neq i} G_{ij}\, u_j^k + \sum_{j\neq i} H_{ij} x_j^k$$

and we may transpose the cost function to give i^{th} term

$$\min_{u_i} \tfrac{1}{2}x_i^T Q_i x_i + \tfrac{1}{2}u_i^T R_i u_i - b_i^T x_i + x_i^T \sum H_{ji}^T \lambda_j^k + u_i^T \sum G_{ji}^T \lambda_j^k \qquad (159)$$

Hence we could proceed to derive the decentralized solution to (147). Our immediate purpose, however, is to verify that this scheme is in fact identical with the NR(K^{-1}) scheme derived above. Recall that we showed this to be the case in the general formulation of Chapter 2. We now demonstrate the identity using the concrete control sequences obtained by each scheme.

(i) *IP Scheme*

The Lagrangian associated with (157) is

$$L(x,u,\lambda) = \tfrac{1}{2}x^T Qx + \tfrac{1}{2}u^T Ru - b^T x + \lambda^{k^T}(Hx+Gu) + \lambda^T(Ax+Bu+Gu^k+Hx^k) \qquad (160)$$

and the necessary conditions of optimality are

$$\frac{\partial L}{\partial \lambda} = Ax + Bu + Gu^k + Hx^k = 0$$

$$\frac{\partial L}{\partial x} = Qx - b + H^T\lambda^k + A^T\lambda = 0 \qquad (161)$$

$$\frac{\partial L}{\partial u} = Ru + G^T\lambda^k + B^T\lambda = 0 .$$

From 161(iii),

$$u = -R^{-1}B^T\lambda - R^{-1}G^T\lambda^k . \qquad (162)$$

Now from 161(ii)

$$\lambda = -A^{T^{-1}}Qx - A^{T^{-1}}H^T\lambda^k + A^{T^{-1}}b \qquad (163)$$

and from 161(i)

$$x = -A^{-1}Bu - A^{-1}Gu^k - A^{-1}Hx^k \quad . \tag{164}$$

Substituting (164) into (163) gives

$$\lambda = A^{T^{-1}}QA^{-1}Bu + A^{T^{-1}}QA^{-1}Gu^k + A^{T^{-1}}QA^{-1}Hx^k - A^{T^{-1}}H^T\lambda^K + A^{T^{-1}}b$$

and this in turn in (162) gives

$$u = -R^{-1}B^TA^{T^{-1}}QA^{-1}Bu - R^{-1}B^TA^{T^{-1}}QA^{-1}Gu^k - R^{-1}B^TA^{T^{-1}}QA^{-1}Hx^k$$
$$+ R^{-1}B^TA^{T^{-1}}H^T\lambda^k - R^{-1}B^TA^{T^{-1}}b - R^{-1}G^T\lambda^k$$

or

$$u = [I + R^{-1}B^TA^{T^{-1}}QA^{-1}B]^{-1}\{-R^{-1}B^TA^{T^{-1}}QA^{-1}Gu^k - R^{-1}B^TA^{T^{-1}}QA^{-1}Hx^k$$
$$+[R^{-1}B^TA^{T^{-1}}H^T - R^{-1}G^T]\lambda^k - R^{-1}B^TA^{T^{-1}}b\}$$

or, removing the factor R^{-1} :

$$u = (R + B^TA^{T^{-1}}QA^{-1}B)^{-1}\{-B^TA^{T^{-1}}QA^{-1}Gu^k - B^TA^{T^{-1}}QA^{-1}Hx^k + (B^TA^{T^{-1}}H-G^T)\lambda^k$$
$$- B^TA^{T^{-1}}b\}. \tag{165}$$

We now substitute in (165) the system state and costate variables

$$x^k = -(A+H)^{-1}(B+G)u^k \tag{166}$$

$$\lambda^k = -(A+H)^{-T}Qx^k + (A+H)^{-T}b \tag{167}$$

to obtain the fundamental control update formula

$$u^{k+1} = (R + B^TA^{T^{-1}}QA^{-1}B)^{-1}\{-B^TA^{T^{-1}}QA^{-1}G + B^TA^{T^{-1}}QA^{-1}H(A+H)^{-1}(B+G)$$
$$+ (B^TA^{T^{-1}}H^T-G^T)(A+H)^{-T}Q(A+H)^{-1}(B+G)]u^k$$
$$+ (B^TA^{T^{-1}}H^T-G^T)(A+H)^{-T}b - B^TA^{T^{-1}}b\} \tag{168}$$

(ii) NR($(R+\hat{W}^TQ\hat{W})^{-1}$) scheme

From (129) we have that

$$u^{k+1} = u^k - (R+B^TA^{-T}QA^{-1}B)^{-1}\, g(u^k)$$

where $g(u^k) = Ru^k + (B+G)^T\lambda^k$

and λ^k, x^k are given by (166), (167) above.

Making successive substitutions we have that

$$u^{k+1} = \{I - (R + B^T A^{-1^T} Q A^{-1} B)^{-1}[R + (B+G)^T (A+H)^{T^{-1}} Q (A+H)^{-1} (B+G))\}u^k$$
$$- (R + B^T A^{-1^T} Q A^{-1} B)^{-1} (B+G)^T (A+H)^{T^{-1}} b$$

and setting $I = K^{-1}K$

$$\text{i.e. } I = (R + B^T A^{-1^T} Q A^{-1} B)^{-1} (R + B^T A^{-1^T} Q A^{-1} B)$$

we find that

$$u^{k+1} = (R + B^T A^{-1^T} Q A^{-1} B)^{-1}\{(R + B^T A^{-1^T} Q A^{-1} B)u^k - [R + (B+G)^T (A+H)^{-1^T}$$
$$Q (A+H)^{-1} (B+G)]u^k - (B+G)^T (A+H)^{-1^T} b.$$

i.e.

$$u^{k+1} = (R + B^T A^{-1^T} Q A^{-1} B)^{-1}\{[B^T A^{-1^T} Q A^{-1} B - (B+G)^T (A+H)^{-1^T} Q (A+H)^{-1} (B+G)]u^k$$
$$- (B+G)^T (A+H)^{-1^T} b\}. \qquad (169)$$

It is a matter of tedious, lengthy algebra to then show that u^{k+1} given by (168) is identical to that given by (169).

§6. Relations between gradient and Lagrangian schemes

We have now reviewed two quite different approaches to the problem of solving the LQ regulator by decomposition and coordination. In the first approach we attack the Lagrange conditions of optimality, and coordinate via multipliers; the update scheme uses a sequence of state vectors (x^k) and/or their duals (λ^k), or a sequence of interconnection vectors (z^k) and/or their duals (μ^k). In the second approach we attack the system gradient and coodinate via gradient information; the update scheme uses a sequence of control vectors (u^k). We now seek to clarify the relationship between these approaches.

At a fundamental level we have already seen in §5.3 that the connection between the two approaches is provided by the costate, or adjoint, equation

$$\lambda = -(A+H)^{-1^T} Qx + (A+H)^{-1^T} b \qquad (116)$$

or, in the 'decoupled case', by the local approximation

$$\hat{\lambda} - -A^{-1^T} Qx + A^{-1^T} b. \qquad (121)$$

But a more immediate question is : does the scheme generated by the NR(K^{-1}) algorithm, where $K = R + \hat{W}^T Q \hat{W}$, have a natural analogue amongst the schemes generated by applying successive approximations to the Lagrange optimality conditions? One suspects that the Takahara scheme is a natural candidate, but this suspicion was first made precise only recently by Cohen [6,7] more than 10 years after the Takahara scheme was published.

Recall that the CM($(R+\hat{W}^T Q \hat{W})^{-1}$) and NR($(R+\hat{W}^T Q \hat{W})^{-1}$) schemes are identical, and equivalent to the IP scheme with coordination by 'reinjection'. In §5.4 this IP scheme was shown to consist of the following problem :

$$\begin{aligned} \min_u \ & \tfrac{1}{2}x^T Qx + \tfrac{1}{2}u^T Ru - b^T x + \lambda^k (Hx + Gu) \\ \text{s.t} \quad & Ax + Bu + z = 0 \\ & z = Gu^k + Hx^k \end{aligned} \qquad (157)$$

where $u^k = -R^{-1}(B+G)^T \lambda^k$.

Now it is a remarkable fact that the Takahara scheme can be derived from a problem which is formally identical to (157), the only difference being the interpretation attached to u^k, x^k and λ^k. If they are considered to be previous iterates (rather than system state and costate), then we can check that (157) indeed generates Takahara.

Write the Lagrangian associated with (157) as

$$L(x,u,\lambda) = \tfrac{1}{2}x^T Qx + \tfrac{1}{2}u^T Ru - b^T x + \lambda^T (Hx+Gu) + \lambda^{k^T}(Ax + Bu + Gu^k + Hx^k) \qquad (170)$$

.....

then differentiate with respect to x,u,λ to obtain necessary conditions of optimality :

$$\frac{\partial \mathcal{L}}{\partial \lambda} = Ax + Bu + Gu^k + Hx^k = 0$$

$$\frac{\partial \mathcal{L}}{\partial x} = Qx + A^T\lambda + H^T\lambda^k = b \qquad (171)$$

$$\frac{\partial \mathcal{L}}{\partial u} = Ru + B^T\lambda + G^T\lambda^k = 0$$

From (171)(iii) we have that

$$u = -R^{-1}B^T\lambda - R^{-1}G^T\lambda^k$$

and substituting this in (i),(ii) gives

$$Ax - BR^{-1}B^T\lambda - BR^{-1}G^T\lambda^k + Gu^k + Hx^k = 0$$

$$Qx + A^T\lambda + H^T\lambda^k = b \qquad (172)$$

and further substituting for u^k gives

$$\begin{bmatrix} A & -BR^{-1}B^T \\ Q & A^T \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} + \begin{bmatrix} H & -BR^{-1}G^T - GR^{-1}(B+G)^T \\ 0 & H^T \end{bmatrix} \begin{bmatrix} x^k \\ \lambda^k \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix}$$

..... (173)

which we recognize as the Takahara scheme(56).

The difference between the two schemes lies in the meaning attached to the given variables x^k, u^k, λ^k. In the gradient scheme they are *global variables*

$$x^k = -(A+H)^{-1}(B+u)u^k$$
$$\lambda^k = -(A+H)^{-1T}Qx^k + (A+H)^{-1T}b \qquad (174)$$

while in the Takahara scheme they are *local variables*

$$x^k = -A^{-1}Bu^k$$
$$\lambda^k = -A^{-1T}Qx^k + A^{-1T}b\,. \qquad (175)$$

Thus, following Cohen, we may interpret the Takahara scheme as a local approximation to the gradient scheme with $K = R + \hat{W}^T Q \hat{W}$. The Takahara scheme carries a lighter computational load - the system gradient never having to be computed - but the price paid is that the scheme is essentially infeasible, and must be computed off-line.

By contrast, the gradient scheme is essentially an on-line computation, and is especially adapted to simulation on the real system, rather than a model of the system, to obtain successive $x^k(t)$, $\lambda^k(t)$ using each $u^k(t)$. This is discussed more extensively in Cohen [6].

Although the defining equations of the Takahara scheme are in terms of $\begin{pmatrix} x \\ \lambda \end{pmatrix}^k$, it is relatively easy to obtain the corresponding update expression for $(u)^k$. It is surprising that this has never been done in the literature on decomposition.

We write the necessary conditions as given in (171):

$$\begin{aligned} Ax + Bu + Gu^k + Hx^k &= 0 \\ Qx + A^T\lambda + H^T\lambda^k &= b \\ Ru + B^T\lambda + G^T\lambda^k &= 0 \end{aligned} \tag{176}$$

where

$$\begin{aligned} \lambda^k &= -A^{-1^T}Qx^k + A^{-1^T}b \\ x^k &= -A^{-1^T}Bu^k \end{aligned}$$

Following through the same sequence as in §5.5 on interaction operators, we have that

$$u = -R^{-1}B^T\lambda - R^{-1}G^T\lambda^k. \tag{177}$$

Now from (176) (ii)

$$\lambda = -A^{T^{-1}}Qx - A^{T^{-1}}H^T\lambda^k + A^{T^{-1}}b \tag{178}$$

and from (176)(i)

$$x = -A^{-1}Bu - A^{-1}Gu^k - A^{-1}Hx^k \ , \tag{179}$$

so that substituting (179) in (178) gives

$$\lambda = +A^{T^{-1}}QA^{-1}Bu + A^{T^{-1}}QA^{-1}Gu^k + A^{T^{-1}}QA^{-1}Hx^k - A^{T^{-1}}H^T\lambda^k + A^{T^{-1}}b$$

and this in turn in (177) gives

$$\begin{aligned} u = -R^{-1}B^TA^{T^{-1}}QA^{-1}Bu - R^{-1}B^TA^{T^{-1}}QA^{-1}Gu^k - R^{-1}B^TA^{T^{-1}}QA^{-1}Hx^k \\ + R^{-1}B^TA^{T^{-1}}H^T\lambda^k - R^{-1}B^TA^{T^{-1}}b - R^{-1}G^T\lambda^k \ . \end{aligned} \tag{180}$$

Finally substituting (175) in (180) gives

$$(I + R^{-1}B^TA^{T^{-1}}QA^{-1}B)u = -R^{-1}B^TA^{T^{-1}}QA^{-1}Gu^k + R^{-1}B^TA^{T^{-1}}QA^{-1}HA^{-1}Bu^k$$
$$+ R^{-1}(B^TA^{T^{-1}}H^T-G^T)A^{T^{-1}}QA^{-1}Bu^k + R^{-1}(B^TA^{T^{-1}}H^T-G^T)A^{T^{-1}}b - R^{-1}B^TA^{T^{-1}}b),$$

and taking out the factor R^{-1}, and writing u^{k+1} for u :

$$u^{k+1} = (R+B^TA^{T^{-1}}QA^{-1}B)^{-1}[B^TA^{T^{-1}}QA^{-1}HA^{-1}B-B^TA^{T^{-1}}QA^{-1}G+(B^TA^{T^{-1}}H^T-G^T)A^{T^{-1}}QA^{-1}B]u^k$$
$$+ [(B^TA^{T^{-1}}H^T - G^T)A^{T^{-1}} - B^TA^{T^{-1}}]b. \qquad (181)$$

Equation (181) is the fundamental expression for the control sequence implicitly generated by the Takahara scheme on the static LQ regulator.

Only in the degenerate, or trivial, case is this control sequence identical to that generated by the gradient scheme, as given by (168) or (169).

It is an open question whether a gradient scheme can be devised that generates the control sequence (181). One suspects strongly that such a scheme would have to involve the 'decoupled' gradient

$$\hat{g}(u) = Ru + B^T\lambda$$

but it has not yet proven possible to isolate such a scheme. Other open questions concern the existence of Lagrangian analogues of the $CG(K^{-1})$ scheme, and gradient analogue of the Lagrangian successive approximations scheme with $K = \begin{bmatrix} A & 0 \\ 0 & A^T \end{bmatrix}$.

CHAPTER 4

Dynamic Interconnected LQ- Regulator : Differential Description.

§1 Introduction and outline of chapter

§2 Formulation of model and derivation of Euler-Lagrange equations

§3 Multilevel solution of Euler-Lagrange equations

§4 Supremal coordination schemes : decomposition based on successive approximations

§5 Successive approximations to differential systems

§6 Proof of convergence of successive approximations scheme

§7 Applicability and numerical computation of successive approximations scheme

§1 Introduction

In this chapter we apply the same schemes outlined in Chapter 3 to the dynamic LQ regulator. The differential description of the LQ regulator is quite standard, but in this case there is the added complexity of a (linear) coupling , or interconnection, function. The system and cost equations are formulated in sufficient generality to encompass all previous models of this system published in the literature.

The plan of the chapter is as follows. In §2. we set up the model , and derive the 2-point boundary problem which constitutes the Euler-Lagrange conditions for optimality. In §3 we systematically attack these conditions, as in Chapter 3, to investigate the possibilities for 2-level decomposition schemes. Again it turns out that some schemes formulated in the abstract in previous published papers fail in this concrete instance. Supremal updating schemes are assumed to be gradient-based, where again the gradient of the system Hamiltonian is easily found.

In §4 an alternative approach is introduced, and the supremal updating scheme is assumed to be of 'reinjection' type. This is generalized as the application of successive approximations to the 2-point boundary value problem that constitutes the Euler-Lagrange conditions. Whereas in Chapter 3 the successive approximations procedure was a straightforward application of Jacobi iteration, the procedure in this chapter is anything but straightforward, and indeed never seems to have been referred to in the literature before. Before tackling this problem, the application of successive approximations to initial value problems and integral equations is reviewed in §5,

and a simple scalar example computed. Then the generalization of the Takahara scheme, namely the iterative method

$$\dot{v}^{k+1}(t) = K(t)v^{k+1}(t) + E(t)v^{k}(t), \quad v_1(t_o) = v_o$$

$$v_2(t_f) = 0$$

is applied to the given 2PBVP with split boundary condition

$$\dot{v} = Cv \quad , \quad v_1(t_o) = v_o, \; v_2(t_f) = 0$$

The scheme is shown to converge over a bounded time interval, the bound being characterized in terms of elements of the transition matrix associated with $K(t)$, and $E(t)$. This generalizes a similar result obtained by Pearson [24].

Finally in §7 the applicability and computational aspects of the successive approximations scheme - both the Takahara special case and the diagonal scheme - are examined.

§2 Formulation of model and derivation of Euler-Lagrange equations

We formulate the interconnected LQ regulator problem in sufficient generality to include all previous models discussed in the literature as special cases. We can then apply known techniques, and formulate new ones, in such a way that connections previously unrecognized will become plain.

We begin with a collection of N subsystems

$$\begin{aligned} \dot{x}_i(t) &= A_i x_i(t) + B_i u_i(t) + z_i(t) \\ x_i(t_o) &= x_{io}, \end{aligned} \tag{1}$$

where $x_i \in \mathbb{R}^{n_i}$ is state vector

$u_i \in \mathbb{R}^{m_i}$ is control vector

and $z_i \in \mathbb{R}^{n_i}$ is interconnection vector,

which are connected together by the 'interconnection constraint ', or 'coupling equation'

$$z_i(t) = \sum G_{ij} u_j(t) + \sum H_{ij} x_j(t). \qquad (2)$$

A typical coupling pattern could be that shown in fig. 1.

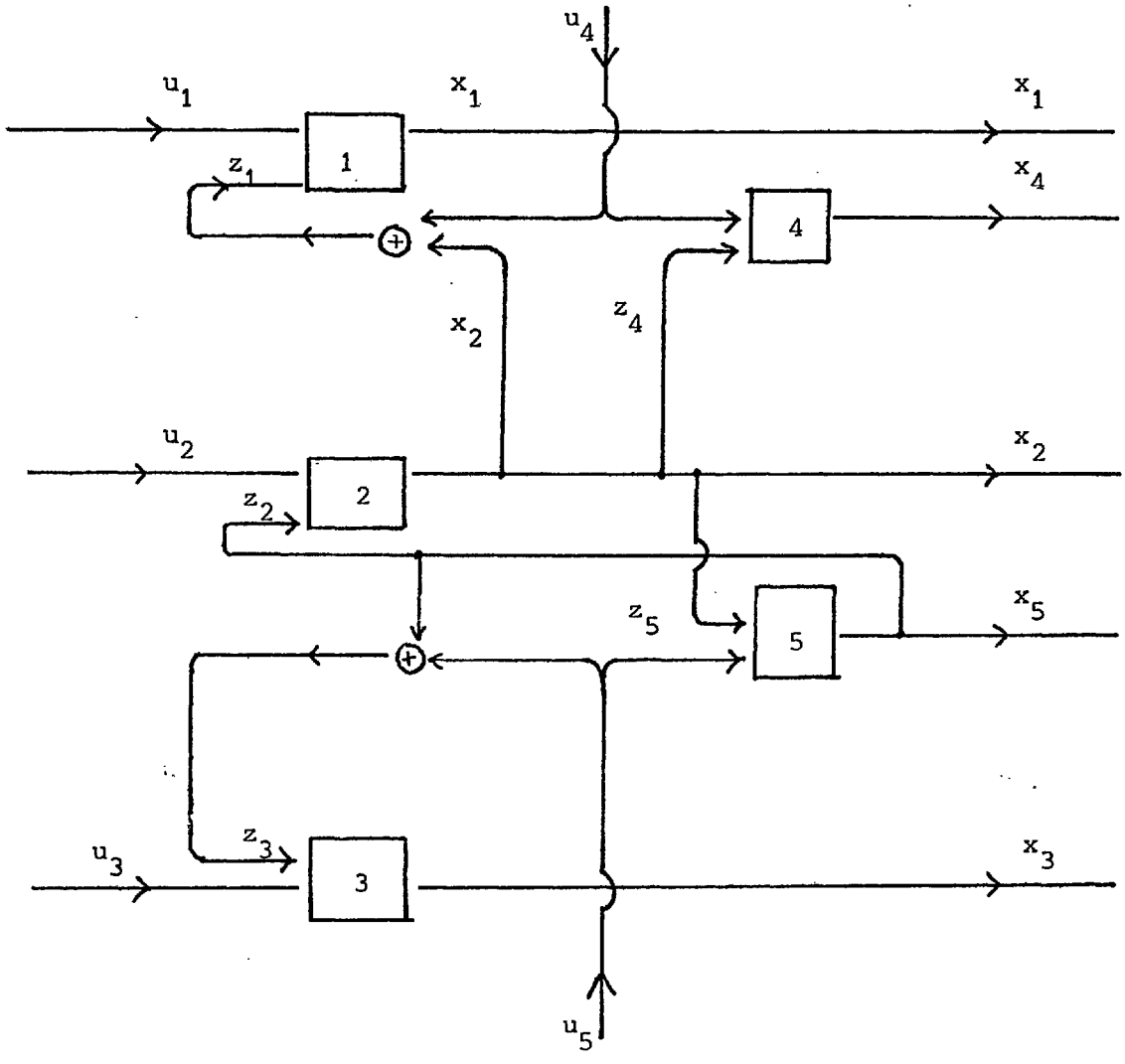


fig. 4.1

Associated with each subsystem is an optimality index of general form

$$C_i = \int_{t_o}^{t_f} \tfrac{1}{2}\|x_i(t)\|^2_{Q_i} + \tfrac{1}{2}\|u_i(t)\|^2_{R_i} + \tfrac{1}{2}\|z_i(t)\|^2_{S_i})\,dt \qquad (3)$$

where $Q_i = Q_i^T \geq 0, \qquad S_i = S_i^T \geq 0$

$$R_i = R_i^T > 0,$$

and we employ the notation $\|v\|^2_W \triangleq v^T W v$.

The addition of the term in z_i is certainly non-standard; and amounts (as we see below) to including cross-coupling terms between u and x in the overall system cost. These terms are needed in one of the decomposition procedures - but in most cases we can ignore the term in z_i, and simply set $S_i = 0$.

If we assume that overall costs are additive in subsystem costs, i,e,

$$\sum C_i = C.$$

(a large assumption, one which rules out much of the interesting conflict between subsystems) then (1) - (3) can be written in aggregate form as

$$\dot{x} = Ax + Bu + z \qquad (4)$$

$$z = Gu + Hx \qquad (5)$$

$$C = \tfrac{1}{2}\int_{t_o}^{t_f} \|x\|^2_Q + \|u(t)\|^2_R + \|z\|^2_S \,dt \qquad (6)$$

Eliminating z we can reduce (4) - (6) to the familiar form of LQ regulator :

$$\dot{x} = (A+H)x + (B+G)u \qquad (7)$$

$$\mathcal{J} = \tfrac{1}{2}\int_{t_o}^{t_f} (x^T(Q+H^TSH)x + u^T(R+G^TSG)u + x^TH^TSGu + u^TG^TSHx)\,dt \qquad (8)$$

or if S = 0,

$$\mathcal{J} = \tfrac{1}{2}\int_{t_o}^{t_f} (x^TQx + u^TRu)dt \qquad (9)$$

Let us solve (7) with cost (9), for this will motivate subsequent discussions. We formulate the system Hamiltonian

$$H(x,u,\lambda,t) = \tfrac{1}{2}x^TQx + \tfrac{1}{2}u^TRu + \lambda^T[(A+H)x + (B+G)u] \qquad (10)$$

and write down the necessary conditions for optimality

$$\frac{\partial H}{\partial \lambda} = \dot{x} = (A+H)x + (B+G)u,\ x(t_o) = x_o \qquad (11)$$

$$\frac{\partial H}{\partial x} = -\dot{\lambda} = Qx + (A+H)^T\lambda,\quad \lambda(t_f) = 0 \qquad (12)$$

$$\frac{\partial H}{\partial u} = 0 = Ru + (B+G)^T\lambda \qquad (13)$$

which results in the 2PBVP

$$\begin{aligned} \dot{x} &= (A+H)x - (B+G)R^{-1}(B+G)^T\lambda \quad , \quad x(t_o) = x_o \\ -\dot{\lambda} &= Qx + (A+H)^T\lambda \qquad\qquad\qquad\quad \lambda(t_f) = 0 \end{aligned} \qquad (14)$$

This can be solved by the usual Riccati procedure.

$$\begin{aligned} &\text{Set} \quad \lambda(t) = P(t)x(t) \\ &\text{then} \quad \dot{\lambda}(t) = P(t)\dot{x}(t) + \dot{P}(t)x(t) \end{aligned} \qquad (15)$$

i.e. $0 = Qx + (A+H)^TPx + P(A+H)x - P(B+G)R^{-1}(B+G)^TPx + \dot{P}x.$

This is satisfied if

$$-\dot{P} = P(A+H) + (A+H)^TP - P(B+G)R^{-1}(B+G)^TP + Q,\quad P(t_f) = 0 \qquad (16)$$

If (16) is solved we then have the optimal control in feedback form as

$$\hat{u} = -R^{-1}(B+G)^TPx$$

and then the optimal cost

$$J^* = x_o^TP(t_o)x_o.$$

The motivation for introducing decomposition and multilevel procedures is to avoid solving (16) when the problem dimensions become very high. In practice, (16) has been solved only for problems of less than 100 dimensions - but we would like to be able to apply the optimal control theory to problems of 1000's of dimensions.

§3. Multilevel solution of Euler-Lagrange equations

Let us attack the problem

$$\min_u \tfrac{1}{2}\int_{t_o}^{t_f} (x^T Q x + u^T R u + z^T S z)\,dt$$

$$\text{s.t.} \qquad \dot{x} = Ax + Bu + z$$

$$z = Gu + Hx$$

directly. The constrained Hamiltonian is

$$H(x,u,\lambda,t) = \tfrac{1}{2}x^T Q x + \tfrac{1}{2}u^T R u + \tfrac{1}{2}z^T S z + \lambda(t)^T(Ax+Bu+z)$$

$$\text{s.t.} \qquad z = Gu + Hx,$$

and this can be *augmented* to

$$\bar{H}(x,u,\lambda,t,\mu) = \tfrac{1}{2}x^T Q x + \tfrac{1}{2}u^T R u + \tfrac{1}{2}z^T S z + \lambda(t)^T(Ax+Bu+z) + \mu(t)^T(z-Gu-Hx).$$

The necessary conditions for optimality are then :

$$\frac{\partial \bar{H}}{\partial \lambda} = \dot{x} = Ax + Bu + z \quad , \quad x(t_o) = x_o \qquad \text{(i)}$$

$$\frac{\partial \bar{H}}{\partial x} = -\dot{\lambda} = Qx + A^T\lambda - H^T\mu \quad , \quad \lambda(t_f) = 0 \qquad \text{(ii)}$$

$$\frac{\partial \bar{H}}{\partial u} = 0 = Ru + B^T\lambda - G^T\mu \qquad \text{(iii)} \qquad (17)$$

$$\frac{\partial \bar{H}}{\partial z} = 0 = Sz + \lambda + \mu \qquad \text{(iv)}$$

$$\frac{\partial \bar{H}}{\partial \mu} = 0 = z - Gu - Hx \qquad \text{(v)}$$

First we check that these necessary conditions reduce to the fundamental optimality conditions (14). Setting S=0, then, we have that from (iii)

$$u = -R^{-1}B^T\lambda + R^{-1}G^T\mu.$$

Substituting this in (i), (ii), (iv) and (v) we have the 2PBVP

$$\begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix} = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} + \begin{bmatrix} I & BR^{-1}G^T \\ 0 & H^T \end{bmatrix} \begin{pmatrix} z \\ \mu \end{pmatrix}, \quad \begin{bmatrix} x(t_o) = x_o \\ \lambda(t_f) = 0 \end{bmatrix}$$

..... (18)

and the algebraic expressions

$$\begin{pmatrix} z \\ \mu \end{pmatrix} = \begin{bmatrix} H & -GR^{-1}(B+G)^T \\ 0 & -I \end{bmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix}. \qquad (19)$$

Note that if we substitute (19) into (18) we get

$$\begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix} = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} + \begin{bmatrix} I & BR^{-1}G^T \\ 0 & H^T \end{bmatrix} \begin{bmatrix} H & -GR^{-1}(B+G)^T \\ 0 & -I \end{bmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix}, \text{ so}$$

$$\begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix} = \left[\begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} + \begin{bmatrix} H & -GR^{-1}(B+G)^T - BR^{-1}G^T \\ 0 & -H^T \end{bmatrix} \right] \begin{pmatrix} x \\ \lambda \end{pmatrix}, \text{ or}$$

$$\begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix} = \begin{bmatrix} A+H & -(B+G)R^{-1}(B+G)^T \\ -Q & -(A+H)^T \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix}, \quad \begin{bmatrix} x(t_o) = x_o \\ \lambda(t_f) = 0 \end{bmatrix}$$

which is (14) again. Hence (18), (19) is a genuinely more 'structured' expression of the optimality conditions (14).

As in the static case above, we wish to solve the set of necessary conditions (17) by a multilevel procedure. Our choice of coordinating variables is more constrained this time : the differential equations in x and λ *must* be solved at the infimal level if anything is to be gained by decomposition, so this rules out x and λ as coordinating variables straight away. Further, as in the static case, the presence of the equation

$$\frac{\partial H}{\partial \mu} = z - Gu - Hx = 0$$

at the infimal level would destroy the possibility of decomposition. Hence there are in fact only two practical possibilities : the use of μ alone as coordinator, or $\begin{pmatrix} \mu \\ z \end{pmatrix}$. We consider these two.

(i) goal coordination through μ alone

In this scheme μ is set at the supremal level, and the equationsto be solved at infimal level are :

$$\text{given } \bar{\mu} : \quad \begin{aligned} \dot{x} &= Ax + Bu + z \\ \dot{\lambda} &= -Qx - A^T\lambda + H^T\bar{\mu} \\ u &= -R^{-1}B^T\lambda + R^{-1}G^T\bar{\mu} \\ z &= -S^{-1}(\lambda+\bar{\mu}) \end{aligned} \tag{20}$$

$$\text{i.e.} \quad \begin{aligned} \dot{x} &= Ax - (BR^{-1}B^T+S^{-1})\lambda + (BR^{-1}G^T-S^{-1})\bar{\mu}, \quad x(t_o) = x_o \\ \dot{\lambda} &= -Qx - A^T\lambda + H^T\bar{\mu}, \quad \lambda(t_f) = 0 \end{aligned} \tag{21}$$

The 2PBVP (21) splits into the N problems :

$$\dot{x}_i = A_i x_i - (B_i R_i^{-1} B_i^T + S_i^{-1})\lambda_i + B_i R_i^{-1} \sum G_{ji}^T \bar{\mu}_j - S_i^{-1}\bar{\mu}_i,$$

$$x_i(t_o) = x_{io};$$

$$\dot{\lambda}_i = -Q_i x_i - A_i^T \lambda_i + \sum H_{ji}^T \bar{\mu}_j, \quad \lambda_i(t_f) = 0,$$

which can be solved independently.

Note that if S=0, there is no expression (20)(iv) in z to substitute in (20)(i) so the system would remain indeterminate. Hence the necessity for a quadratic term in z for this method.

At the supremal level μ could be updated by a simple gradient procedure.

Let $e = H_\mu = z - Gu - Hx$.

Then the stee pest ascent algorithm could require the update :

$$\mu^{k+1} = \mu^k + \varepsilon e^k .$$

Alternatively we could use the conjugate gradient algorithm.

The scheme may be envisaged as follows :

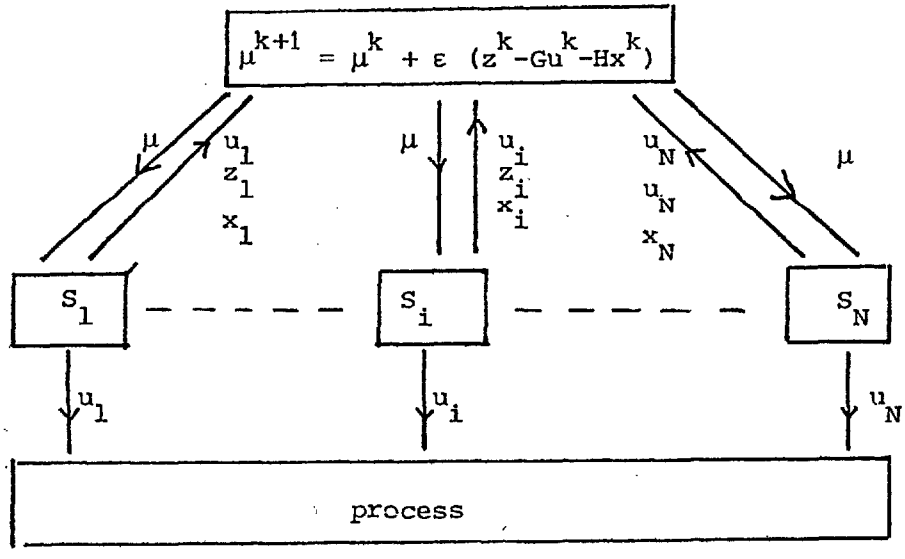


fig. 4.2

(ii) primal/dual coordination through $\begin{pmatrix}\mu\\ z\end{pmatrix}$.

In this scheme $\begin{pmatrix}\mu\\ z\end{pmatrix}$ is set at the supremal level, leaving 3 equations to be solved at the infimal level :

$$\dot{x} = Ax + Bu + \bar{z}$$
$$\dot{\lambda} = -Qx - A^T\lambda + H^T\bar{\mu}$$
$$u = -R^{-1}B^T\lambda + R^{-1}G^T\bar{\mu}$$

i.e.
$$\dot{x} = Ax - BR^{-1}B^T\lambda + BR^{-1}G^T\bar{\mu} + \bar{z}, \qquad x(t_o) = x_o$$
$$\dot{\lambda} = -Qx - A^T\lambda + H^T\bar{\mu}\ , \qquad \lambda(t_f) = 0 \qquad (22)$$

The 2PBVP (22) splits into the N problems

$$\dot{x}_i = A_i x_i - B_i R_i^{-1} B_i^T \lambda_i + B_i R_i^{-1} \sum G_{ji}^T \bar{\mu}_j + \bar{z}_i, \qquad x_i(t_o) = x_{io}$$

$$\dot{\lambda}_i = -Q_i x_i - A_i^T \lambda_i + \sum H_{ji}^T \bar{\mu}_j, \qquad \lambda_i(t_f) = 0$$

and these can be solved independently.

At the supremal level $\begin{pmatrix}\mu\\ z\end{pmatrix}$ could be updated by a simple gradient procedure .

Let $e = \begin{pmatrix} H_\mu \\ H_z \end{pmatrix} = \begin{pmatrix} z - Gu - Hx \\ Sz + \lambda + \mu \end{pmatrix}$

Then the steepest descent algorithm would require the update :

$$\begin{pmatrix}\mu\\ z\end{pmatrix}^{k+1} = \begin{pmatrix}\mu\\ z\end{pmatrix}^{k} + \varepsilon\, e^{k}.$$

(iii) The only other possible coordination variables are z and u. Again some authors have proposed them as coordinating variables without sufficient regard to the real possibilities of decomposition. For instance Grataloup and Titli [37] propose the use of z as coordinating variable, calling this 'model coordination'. The equations to be solved at the infimal level are then :

$$\left.\begin{aligned} \dot{x} &= Ax + Bu + \bar{z} \\ \dot{\lambda} &= -Qx - A^T\lambda + H^T\mu \end{aligned}\right\} \qquad (23)$$

$$\left.\begin{aligned} G^T\mu &= Ru + B^T\lambda \\ Gu &= -Hx + \bar{z} \end{aligned}\right\} \qquad (24)$$

Again it is clear that the linear system (24) cannot be solved in decomposed fashion. Breaking it into N subproblems :

$$\left.\begin{aligned} \sum_{j=1}^{N} G_{ji}^T \mu_j &= R_i u_i + B_i^T \lambda_i \\ \sum_{j=1}^{N} G_{ij} u_j &= -\sum_{\substack{j=1 \\ j\neq i}}^{N} H_{ij} x_j + \bar{z}_i \end{aligned}\right\} \qquad (25)$$

we see that there is no μ_i nor u_i on the LHS. Even if u_i is solved by 25(i), this still leaves the problem of how the i^{th} subsystem is to know μ_j, $j = 1$, unless specifically informed by coordinator, and 25(ii) offers no solution at all for μ_i.

In spite of the decade of intense discussion of decomposition procedures the derivation of the simple update formulae for the LQ regulator has never been attempted before in the literature. Hence problems such as that discussed above have not come to light.

§3 Successive approximations iteration on Euler-Lagrange equations

We now return to the Euler-Lagrange equations (14) :

$$\dot{x} = (A+H)x - (B+G)R^{-1}(B+G)^T \lambda, \qquad x(t_o) = x_o$$

$$\dot{\lambda} = -Qx - (A+H)^T \lambda, \qquad \lambda(t_f) = 0$$

and consider the use of successive approximations.

Our approach will be to find suitable decompositions of the Euler-Lagrange equations. The most obvious procedure, first formulated by Takahara [3] in 1964, is to propose a supremal level updating procedure (based on the algebraic expression (19)) :

$$\begin{bmatrix} z \\ \mu \end{bmatrix}^{k+1} = \begin{bmatrix} H & -GR^{-1}(B+G)^T \\ 0 & -I \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix}^k \qquad \ldots\ldots \quad (26)$$

which is a simple algebraic substitution, and an infimal level problem based on the 2PBVP (18) :

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} + \begin{bmatrix} I & BR^{-1}G^T \\ 0 & H^T \end{bmatrix} \begin{bmatrix} \bar{z} \\ \bar{\mu} \end{bmatrix}, \quad \begin{matrix} x(t_o) = x_o \\ \lambda(t_f) = 0 \end{matrix}$$

$\ldots\ldots \quad (27)$

where $(\bar{z},\bar{\mu})$ is given by supremal level.

The point about this procedure is that (27) can be decomposed into N 2PBVP's :

$$i=1..N \quad \begin{cases} \dot{x}_i = A_i x_i - B_i R_i^{-1} B_i^T \lambda_i + \bar{z}_i + B_i R_i^{-1} \sum G_{ji}^T \bar{\mu}_j, & x_i(t_o) = x_{io} \\ \dot{\lambda}_i = -Q_i x_i - A_i^T \lambda_i + \sum H_{ji}^T \bar{\mu}_j, & \lambda_i(t_f) = 0 . \end{cases}$$

Analytically, by substituting (26) into (27) the procedure is seen to be

$$\begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix}^{k+1} = \begin{pmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix}^{k+1} + \begin{pmatrix} H & -GR^{-1}(B+G)^T - BR^{-1}G^T \\ 0 & -H^T \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix}^{k},$$

$$\begin{pmatrix} x(t_o) = x_o \\ \lambda(t_f) = 0 \end{pmatrix} \qquad \ldots\ldots \quad (28)$$

which constitutes the stationarity condition for the problem

$$\left.\begin{aligned} \min_u J &= \int \{ \tfrac{1}{2}x^T Q x + \tfrac{1}{2}u^T R u + \lambda^{k^T}(Hx + Gu)\} dt \\ \text{s.t.} \quad \dot{x} &= Ax + Bu + z \\ z &= Gu^k + Hx^k \end{aligned}\right\} \qquad (29)$$

when $u^k = -R^{-1}(B+G)^T \lambda^k$.

We shall use this formulation of the Takahara scheme in Chapter 5 §4 below when we show that use of Mesarovic's interaction operators leads to a scheme which is formally identical. The difference lies in the significance of the x^k and λ^k : in the IP formulation they are *system*, or *global*, states and costates, whereas in the Takahara scheme they are *local* approximations. This is explained in detail below.

Before tackling the question of convergence, let us note that (23) suggests a general distributive procedure for solving the Euler-Lagrange equations, of which the Takahara scheme will be a special case. Putting ω in place of $\begin{pmatrix} x \\ \lambda \end{pmatrix}$, we see that (28) has the form

$$\dot{\omega}^{k+1} = K\omega^{k+1} + E\omega^k \qquad (30)$$

with split boundary conditions, whereas the aggregate necessary condition (14) has the form

$$\dot{\omega} = C\omega, \quad C = K+E$$

In other words, the Takahara scheme is a successive approximation solution of the aggregate necessary conditions (14) in exact analogy with the Jacobi procedure for algebraic linear equations of the form

$$b = Cv$$

where C is replaced by K+E, and the iterative expression

$$v^{k+1} = Kv^{k+1} + Ev^{k}$$

is solved.

However the more general formulation (30) suggests a range of decomposition procedures which have not previously been mentioned in the literature. One very interesting case is the following

(ii) $$K = \begin{pmatrix} A & 0 \\ 0 & -A^T \end{pmatrix}, \qquad E = \begin{pmatrix} H & -(B+G)R^{-1}(B+G)^T \\ -Q & -H^T \end{pmatrix} \tag{31}$$

This decouples the 2PBVP with split boundary conditions into 2 independent equations, simplifying solution procedures enormously. Writing (31) in full

$$\left.\begin{aligned} \dot{x} &= Ax + Hx^k - (B+G)R^{-1}(B+G)^T\lambda^k, \quad x(t_o) = x_o \\ \dot{\lambda} &= -A^T\lambda - Qx^k - H^T\lambda^k \qquad\qquad , \quad \lambda(t_f) = 0 \end{aligned}\right\} \tag{32}$$

we see that it is equivalent to the N subsystems

$$i=1\ldots N \qquad \dot{x}_i = A_i x_i + \sum_j H_{ij} x_j^k - B_i R_i^{-1} B_i^T \lambda_i^k - B_i R_i^{-1} \sum_j G_{ij}^T \lambda_j^k \tag{33}$$

$$- \sum_j G_{ij} R_j^{-1} B_j \lambda_j^k - \sum_j G_{ij} R_j^{-1} \sum_\ell G_{\ell j}^T \lambda_\ell^k, \quad x_i(t_o) = x_i$$

$$\dot{\lambda}_i = -A_i\lambda_i - Q_i x_i^k - \sum H_{ij}^T \lambda_j^k, \qquad \lambda_i(t_f) = 0 \tag{34}$$

This appears to be a very satisfactory decentralised procedure for solving the interconnected LQ regulator, *on-line*. Each subsystem solves its own necessary conditions for optimality, using past outputs obtained from other subsystem interconnections.

Further artificial procedures are possible. Consider the case

$$\text{(iii)} \quad K = I,\ E = \begin{pmatrix} A+H-I & -(B+G)R^{-1}(B+G)^T \\ -Q & -(A+H)^T-I \end{pmatrix} \tag{36}$$

which written in full gives

$$\dot{x} = x + (A+H-I)x^k - (B+G)R^{-1}(B+G)^T\lambda^k,\ x(t_o) = x_o$$
$$\dot{\lambda} = \lambda - Qx^k - [(A+H)^T+I]\lambda^k,\ \lambda(t_f) = 0 \tag{37}$$

and this again can be decoupled into N subsystems. Scheme (iii) might be parameterized to give potentially faster convergence, by setting

$$\text{(iv)} \quad K = \gamma I,\ E = \begin{pmatrix} A+H-\gamma I & -(B+G)R^{-1}(B+G)^T \\ -Q & -(A+H)^T-\gamma I \end{pmatrix} \tag{38}$$

We now proceed to discuss several aspects of this general procedure. We ask :

(1) Does it converge?

(2) If it converges, does it give the optimal u*?

(3) How is the scheme treated numerically?

(1) Convergence

No general proof of the convergence of the Takahara scheme exists in the literature. Takahara himself [31]derived an extremely complicated convergence condition, for the case of stationary matrices A,B,Q,R, using Laplace transformation theory. Later Pearson [24] demonstrated convergence of the scheme without conditions over some interval $[t_o,\tau]$, $\tau \leq t_f$. Cohen et al [41] later adopted an indirect approach by computing the difference between successive values of the cost

$$\Delta J_{k+1} = J_{k+1} - J_k. \tag{39}$$

They derived a condition under which this difference is positive-definite, hence ensuring that the cost monotonically decreases.

This turns out to be the same condition governing convergence of the $NR(K^{-1})/CM(K^{-1})$ gradient schemes discussed below in Chapter 5. The condition suggests that the algorithm will converge if R is sufficiently large; Singh and Hassan [40] claim to have found a numerical case of divergence with R sufficiently small.

We shall adopt a new approach, utilizing the generalization of the Takahara scheme, viz

$$\dot{\omega}^{k+1} = K\omega^{k+1} + E\omega^{k} \quad , \quad \begin{pmatrix} \omega_1(t_o) = \omega_o \\ \omega_2(t_f) = 0 \end{pmatrix} \tag{40}$$

Like Pearson, we prove that the scheme converges over $[t_o, \tau)$, $\tau < t_f$, and characterize the upper bound on the time interval over which the scheme converges.

Before introducing the theorem, we review the application of successive approximations to differential equations.

§5 Successive approximations to differential systems

The method of successive approximations as applied to the initial value problem is well known (Mikhlin and Smolitskiy [59], Riesz and Sz-Nagy [55]). Consider the problem

$$\dot{x} = f(x,t), \quad x(t_o) = x_o \tag{41}$$

over the interval $[t_o, t_1]$. It is known that if f is sufficiently smooth, namely that it satisfies a Lipschitz condition

$$|f(x_1,t) - f(x_2,t)| \leq K|x_1 - x_2| \ , \tag{42}$$

then the successive approximations (Picard iteration)

$$x^{k+1}(t) = x_o + \int_{t_o}^{t} f(x^k(\tau),\tau)d\tau \tag{43}$$

converges uniformly; and its limit is the unique solution of the initial value problem (41). The initial approximation may be selected arbitrarily,

e.g. $x^{o}(t) = x_o, \forall t.$

Reformulating (41) as an integral equation

$$x(t) = x_o + \int_{t_o}^{t} f(\tau, x(\tau))d\tau \tag{44}$$

we see how the iteration (43) arises naturally.

Now consider the linear differential system

$$\dot{x} = A(t)x(t), \quad x(t_o) = x_o \tag{45}$$

over the interval $[t_o, t_f]$. Let us split the matrix A(t) into 2 parts, such that

$$A(t) = K(t) + E(t) \tag{46}$$

then (45) may be written as

$$\dot{x}(t) = K(t)x(t) + E(t)x(t), \quad x(t_o) = x_o \tag{47}$$

Using the Variation of Constants formula, we may solve (47) to obtain

$$x(t) = \Phi(t,t_o)x_o + \int_{t_o}^{t} \Phi(t,\tau)E(\tau)x(\tau)d\tau \tag{48}$$

where $\frac{d}{dt}\Phi(t,t_o) = K(t)\Phi(t,t_o)$

and $\Phi(t_o,t_o) = \Phi(t,t) = I$.

Let us write (48) in the notation of Mikhlin and Smolitskiy as

$$\phi(t) = f(t) + \lambda \int_{t_o}^{t} \kappa(t,\tau)\phi(\tau)d\tau \tag{49}$$

or $$\phi = f + \lambda \kappa \phi, \tag{50}$$

where $f(t) = \Phi(t,t_o)x_o$

$\kappa(t,\tau) = \Phi(t,\tau)E(\tau)$

and $\lambda = 1$.

If successive approximations are applied to (50) we have

$$\phi^{k+1} = f + \lambda \kappa \phi^{k} \tag{51}$$

It is well-known [Mikhlin and Smolitskiy, p 282] that the iteration process (51) converges to the solution for all λ, provided the interval (t_o, t_f) is finite and the kernel $\kappa(t,\tau)$ is Volterra.

Now the Fredholm equation (49) is said to be a Volterra equation if

$$\kappa(t,\tau) = 0 \quad \text{for } t_o \leq t \leq \tau \leq t_f. \tag{52}$$

In this case, then,

$$\begin{aligned} \kappa\phi &= \int_{t_o}^{t_f} \kappa(t,\tau)\phi(\tau)d\tau \\ &= \int_{t_o}^{t} \kappa(t,\tau)\phi(\tau)d\tau + \int_{t}^{t_f} \kappa(t,\tau)\phi(\tau)d\tau \\ &= \int_{t_o}^{t} \kappa(t,\tau)\phi(\tau)d\tau, \end{aligned} \tag{53}$$

by (52).

When we examine (49) we see that the integral is indeed in Volterra form, and so we know that the iteration (51) converges to the solution. Transcribing this back to our original notation, we know that the iteration

$$x^{k+1}(t) = \Phi(t,t_o)x_o + \int_{t_o}^{t} \Phi(t,\tau)E(\tau)x^k(\tau)d\tau \tag{54}$$

or equivalently, the iteration based on the solution to the differential equation

$$\dot{x}^{k+1}(t) = K(t)x^{k+1}(t) + E(t)x^k(t), \; x^{k+1}(t_o) = x_o, \tag{55}$$

converges to the solution.

We now illustrate these ideas on a simple scalar example.

Example

$$\dot{x} = -2x, \; x(0) = 1, \quad 0 \leq t \leq 1 \; . \tag{56}$$

In this case the solution is

$$x(t) = e^{-2t} \; , \tag{57}$$

so that

$$x(0) = 1.000$$

$$x(.2) = 0.670$$

$$x(.4) = 0.449$$

$$x(.6) = 0.302$$

$$x(.8) = 0.202$$

$$x(1.0) = 0.135 \; .$$

This curve is shown in fig 4.3



fig 4.3

Applying the successive approximations scheme (55), we have

$$\dot{x}^{k+1} = -x^{k+1} - x^{k}, \; x^{k+1}(0) = 1.$$

Solving this system we have by the Variation of Constants formula

$$x^{k+1}(t) = e^{-t} - \int_0^t e^{-(t-\tau)} x^k(\tau) d\tau$$

i.e.

$$\boxed{x^{k+1}(t) = e^{-t} - e^{-t} \int_0^t e^{\tau} x^k(\tau) d\tau} \tag{58}$$

We set $x^0(t) = 1 \quad \forall t$

then

$$x^1(t) = e^{-t} - e^{-t} \int^t e^{\tau} d\tau$$

$$= e^{-t} - e^{-t} (e^t - 1)$$

$$= e^{-t} - e^{t} + 2 \; .$$

$$x^2(t) = -e^{-t} - e^{-t}\int_0^t e^{\tau}\, e^{-\tau}(-e^{\tau}+2)\,d\tau$$

$$= e^{-t} + e^{-t}\int_0^t (e^{\tau}-2)\,d\tau$$

$$= e^{-t} + e^{-t}(e^{t}-1-2t)$$

$$= e^{-t}(e^{t}-2t).$$

$$x^3(t) = e^{-t} - e^{-t}\int_0^t e^{\tau}e^{-\tau}(e^{\tau}-2\tau)\,d\tau$$

$$= e^{-t} - e^{-t}\int_0^t (e^{\tau}-2\tau)\,d\tau$$

$$= e^{-t} - e^{-t}(e^{t}-1-t^2)$$

$$= e^{-t}(-e^{t}+2+t^2).$$

$$x^4(t) = e^{-t} - e^{-t}\int_0^t e^{\tau}\, e^{-\tau}(-e^{\tau}+2+\tau^2)\,d\tau$$

$$= e^{-t} - e^{-t}\int_0^t (-e^{\tau}+2+\tau^2)\,d\tau$$

$$= e^{-t} - e^{-t}(-e^{t}+1+2t+\frac{1}{3}t^3)$$

$$= e^{-t}(e^{t}-2t-\frac{2}{3!}t^3).$$

The solution pattern is by now clear. We have the alternating sequence :

$$x^0(t) = 1 = e^{-t}(e^{t}) = e^{-t}(1+t+\frac{t^2}{2!}+\frac{t^3}{3!}+\frac{t^4}{4!}+\frac{t^5}{5!}+\dots)$$

$$x^2(t) = e^{-t}(e^{t}-2t) = e^{-t}(1-t+\frac{t^2}{2!}+\frac{t^3}{3!}+\frac{t^4}{4!}+\frac{t^5}{5!}+\dots)$$

$$x^4(t) = e^{-t}(e^{t}-2t-\frac{2}{3!}t^3) = e^{-t}(1-t+\frac{t^2}{2!}-\frac{t^3}{3!}+\frac{t^4}{4!}+\frac{t^5}{5!}+\dots)$$

$$x^6(t) = e^{-t}(e^{t}-2t-\frac{2}{3!}t^3-\frac{2}{5!}t^5) = e^{-t}(1-t+\frac{t^2}{2!}-\frac{t^3}{3!}+\frac{t^4}{4!}-\frac{t^5}{5!}+\dots)$$

.
.
.

and

$$x^1(t) = e^{-t}(-e^t+2) \quad = e^{-t}(1-t - \frac{t^2}{2!} - \frac{t^3}{3!} - \frac{t^4}{4!} - \frac{t^5}{5!} - \ldots.)$$

$$x^3(t) = e^{-t}(-e^t+2+t^2) = e^{-t}(1-t+ \frac{t^2}{2!} - \frac{t^3}{3!} - \frac{t^4}{4!} - \frac{t^5}{5!} - \ldots.)$$

$$x^5(t) = e^{-t}(-e^t+2+ \frac{2t^2}{2!} + \frac{2t^4}{4!}) = e^{-t}(1-t+ \frac{t^2}{2!} - \frac{t^3}{3!} + \frac{t^4}{4!} - \frac{t^5}{5!} + \ldots.).$$

Both sequences are clearly converging to the solution $x^*(t)$ for

$$x^*(t) = e^{-2t} = e^{-t}(1-t+ \frac{t^2}{2} - \frac{t^3}{3!} + \frac{t^4}{4!} - \frac{t^5}{5!} + \frac{t^6}{6!} - \frac{t^7}{7!} + \ldots.).$$

In particular, the even terms are bounded *below* by the solution, as successive alternate terms of e^t are replaced by their negative counterparts; and the odd terms are bounded *above* by the solution, as successive alternative terms of $-e^t$ are replaced by their positive counterparts.

This sequence was computed on the CDC 6400 at Imperial College, with the following results. (to 3 significant figures).

$t =$	0	0.2	0.4	0.6	0.8	1.0
$x^o(t) =$	1.000	1.000	1.000	1.000	1.000	1.000
$x^1(t) =$	1.000	0.637	0.341	0.098	-0.101	-0.264
$x^2(t) =$	1.000	0.673	0.464	0.341	0.281	0.264
$x^3(t) =$	1.000	0.670	0.448	0.295	0.186	0.104
$x^4(t) =$	1.000	0.670	0.449	0.302	0.204	0.142
$x^5(t) =$	1.000	0.670	0.449	0.301	0.202	0.134
$x^6(t) =$	1.000	0.670	0.449	0.301	0.202	0.134
cf $x^*(t) = e^{-2t} =$	1.000	0.670	0.449	0.301	0.202	0.135

It may be seen that even terms $x^0(t)$, $x^2(t)$,... converge *down* to the solution, while odd terms $x^1(t)$, $x^3(t)$... converge *up* to the solution - but convergence is almost complete by the 5^{th} iteration. ‖

§6 Proof of convergence of successive approximations scheme

We now proceed to the main contribution of this chapter, viz the discussion of the 2PBVP

$$\begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix} = C \begin{pmatrix} x \\ \lambda \end{pmatrix}, \qquad \begin{matrix} x(t_o) = x_0 \\ \lambda(t_f) = 0 \end{matrix}. \tag{59}$$

We consider the novel successive approximations scheme

$$\begin{pmatrix} \dot{x} \\ \lambda \end{pmatrix}^{k+1} = K \begin{pmatrix} x \\ \lambda \end{pmatrix}^{k+1} + E \begin{pmatrix} x \\ \lambda \end{pmatrix}^{k}, \qquad \begin{matrix} x^{k+1}(t_o) = x_o \\ \lambda^{k+1}(t_f) = 0 \end{matrix} \tag{60}$$

where $K+E = C$. It is not immediately apparent that this scheme can be put into the form of a contraction map but this we shall proceed to show.

First we derive an expression for $\begin{pmatrix} x \\ \lambda \end{pmatrix}^{k+1}$ in terms of $\begin{pmatrix} x \\ \lambda \end{pmatrix}^{k}$: a task that is complicated by the split boundary conditions.

For the general scheme (60) we have, by the Variation of Constants formula,

$$\begin{pmatrix} x(t) \\ \lambda(t) \end{pmatrix}^{k+1} = \Phi(t,t_o) \begin{pmatrix} x_o \\ \lambda(t_o) \end{pmatrix} + \int_{t_o}^{t} \Phi(t,\tau) E \begin{pmatrix} x(\tau) \\ \lambda(\tau) \end{pmatrix}^{k} d\tau , \tag{61}$$

where $\dfrac{d}{dt}\Phi(t,t_o) = K(t)\Phi(t,t_o)$

and $\Phi(t_o,t_o) = \Phi(t,t) = I$.

Setting $\Phi(t_f,\tau) = \begin{pmatrix} \phi_{11}^{\tau} & \phi_{12}^{\tau} \\ \phi_{21}^{\tau} & \phi_{22}^{\tau} \end{pmatrix}$ (62)

and $\Phi(t_f,t_o) = \begin{bmatrix} \phi^o_{11} & \phi^o_{12} \\ \phi^o_{21} & \phi^o_{22} \end{bmatrix}$, $\Phi(t,t_o) = \begin{bmatrix} {}^t\phi^o_{11} & {}^t\phi^o_{12} \\ {}^t\phi^o_{21} & {}^t\phi^o_{22} \end{bmatrix}$ (63)

we have at terminal time t_f, from (61)

$$0 = \lambda^{k+1}(t_f) = \phi^o_{21}x_o + \phi^o_{22}\lambda^{k+1}(t_o) + \int_{t_o}^{t_f} [\phi^\tau_{21} \quad \phi^\tau_{22}]E \begin{bmatrix} x(\tau) \\ \lambda(\tau) \end{bmatrix}^k d\tau ;$$

Hence, since ϕ^o_{22} is invertible as $\Phi(t_f,t_o)$ is invertible, we have that

$$\lambda^{k+1}(t_o) = -(\phi^o_{22})^{-1}\phi^o_{21}x_o - (\phi^o_{22})^{-1}\int_{t_o}^{t_f} [\phi^\tau_{21} \quad \phi^\tau_{22}]E \begin{bmatrix} x(\tau) \\ \lambda(\tau) \end{bmatrix}^k d\tau. \quad \ldots\ldots (64)$$

Substituting for $\lambda^{k+1}(t_o)$ in (61) now gives

$$\begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix}^{k+1} = \begin{bmatrix} {}^t\phi^o_{11} & {}^t\phi^o_{12} \\ {}^t\phi^o_{21} & {}^t\phi^o_{22} \end{bmatrix} \left\{ \begin{bmatrix} x_o \\ -(\phi^o_{22})^{-1}\phi^o_{21}x_o - (\phi^o_{22})^{-1}\int_{t_o}^{t_f} [\phi^\tau_{21} \quad \phi^\tau_{22}]E \begin{bmatrix} x(\tau) \\ \lambda(\tau) \end{bmatrix}^k d\tau \end{bmatrix} \right.$$

$$\left. + \int_{t_o}^{t} \begin{bmatrix} {}^t\phi^\tau_{11} & {}^t\phi^\tau_{12} \\ {}^t\phi^\tau_{21} & {}^t\phi^\tau_{22} \end{bmatrix} E \begin{bmatrix} x(\tau) \\ \lambda(\tau) \end{bmatrix}^k d\tau \right\},$$

where $\Phi(t,\tau) = \begin{bmatrix} {}^t\phi^\tau_{11} & {}^t\phi^\tau_{12} \\ {}^t\phi^\tau_{21} & {}^t\phi^\tau_{22} \end{bmatrix}$,

so that

$$\begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix}^{k+1} = \begin{bmatrix} {}^t\phi^o_{11} \\ {}^t\phi^o_{21} \end{bmatrix} x_o - \begin{bmatrix} {}^t\phi^o_{12} \\ {}^t\phi^o_{22} \end{bmatrix} \phi^{o\,-1}_{22}\phi^o_{21}x_o - \begin{bmatrix} {}^t\phi^o_{12} \\ {}^t\phi^o_{22} \end{bmatrix} \phi^{o\,-1}_{22} \int_{t_o}^{t_f} [\phi^\tau_{21}\phi^\tau_{22}]E \begin{bmatrix} x(\tau) \\ \lambda(\tau) \end{bmatrix}^k d\tau$$

$$+ \int_{t_o}^{t} \begin{bmatrix} {}^t\phi^\tau_{11} & {}^t\phi^\tau_{12} \\ {}^t\phi^\tau_{21} & {}^t\phi^\tau_{22} \end{bmatrix} E \begin{bmatrix} x(\tau) \\ \lambda(\tau) \end{bmatrix}^k d\tau. \qquad (65)$$

We may write (65) in operator form as

$$z^{k+1} = h + Pz^k + Qz^k = Rz^k + h \qquad (66)$$

where $Z(t) \triangleq \begin{pmatrix} x(t) \\ \lambda(t) \end{pmatrix}$

$$(Pz)(t) \triangleq -\begin{pmatrix} \phi_{12}(t,t_o) \\ \phi_{22}(t,t_o) \end{pmatrix} \phi_{22}(t,t_o)^{-1} \int_{t_o}^{t_f} [\phi_{21}(t_f,\tau)\phi_{22}(t_f,\tau)]E(\tau) \begin{pmatrix} x(\tau) \\ \lambda(\tau) \end{pmatrix} d\tau$$

$$\text{or } (Pz)(t) = -\int_{t_o}^{t_f} C(t,\tau)z(\tau)d\tau \tag{67}$$

$$\text{and } (Qz)(t) \triangleq \int_{t_o}^{t} \Phi(t,\tau)E(\tau)z(\tau)d\tau$$

$$\text{or } (Qz)(t) = \int_{t_o}^{t} D(t,\tau)z(\tau)d\tau\ , \tag{68}$$

and $$h(t) = \begin{pmatrix} {}^{t}\phi_{11}^{o} \\ {}^{t}\phi_{21}^{o} \end{pmatrix} x_o - \begin{pmatrix} {}^{t}\phi_{21}^{o} \\ {}^{t}\phi_{22}^{o} \end{pmatrix} \phi_{22}^{o\,-1}\ \phi_{21}^{o}\ x_o.$$

These operators are clearly well-defined on the space $C[t_o,t_f]$ with standard norm

$$\|v\|_\tau = \sup_{t_o \le t \le \tau} |z(t)| \tag{69}$$

where $|z(t)|$ is the Euclidean length in $\mathbb{R}^{2n}$ of the vector $z(t)$, where $z(t) = \begin{pmatrix} x(t) \\ \lambda(t) \end{pmatrix}$, $x(t) \in \mathbb{R}^n$, $\lambda(t) \in \mathbb{R}^n$.

We now have the following

Theorem The operator R in (66) is a contraction map, provided the time interval satisfies

$$t_f - t_o < \frac{1}{\sup_{t,\tau} |C(t,\tau)| + \sup_{t,\tau} |D(t,\tau)|} \tag{70}$$

Proof

$$||Rx - Ry|| = \sup_{t_o \le t \le t_f} |Px(t) - Py(t) + Qx(t) - Qy(t)|$$

$$= \sup_t \left| \int_{t_o}^{t_f} -C(t,\tau)\,[(x(\tau) - y(\tau)]\,d\tau + \int_{t_o}^{t} D(t,\sigma)\,[(x(\sigma) - y(\sigma)]\,d\sigma \right|$$

$$\le \sup_t \left\{ \left| \int_{t_o}^{t_f} C(t,\tau)[(x(\tau) - y(\tau)]\,d\tau \right| + \left| \int_{t_o}^{t} D(t,\sigma)[(x(\sigma) - y(\sigma)]\,d\sigma \right| \right\}$$

$$\le \sup_t \left\{ \int_{t_o}^{t_f} |C(t,\tau)|\,|x(\tau) - y(\tau)|\,d\tau + \int_{t_o}^{t} |D(t,\sigma)|\,|x(\sigma) - y(\sigma)|\,d\sigma \right\}$$

$$\le \sup_t \left\{ \int_{t_o}^{t_f} \sup_{t,\tau} |C(t,\tau)| \sup_{\tau \le t_f} |x(\tau) - y(\tau)|\,d\tau + \int_{t_o}^{t} \sup_{t,\sigma} |D(t,\sigma)| \sup_{\sigma \le t} |x(\sigma) - y(\sigma)|\,d\sigma \right\}$$

and setting $\alpha = \sup_{t,\tau} |C(t,\tau)|$, $\beta = \sup_{t,\sigma} |D(t,\sigma)|$,

$$\le \sup_t \left\{ \alpha \int_{t_o}^{t_f} \sup_{\tau \le t_f} |x(\tau) - y(\tau)|\,d\tau + \beta \int_{t_o}^{t} \sup_{\sigma \le t_f} |x(\sigma) - y(\sigma)|\,d\sigma \right\}$$

and setting $||x - y||_{t_f} = \sup_{\tau \le t_f} |x(\tau) - y(\tau)|$,

$$\le ||x - y||_{t_f} \sup_t \left\{ \alpha \int_{t_o}^{t_f} d\tau + \beta \int_{t_o}^{t} d\sigma \right\}$$

$$\le ||x - y||_{t_f} \max_t \left\{ \alpha\,(t_f - t_o) + \beta\,(t - t_o) \right\}$$

$$= ||x - y||_{t_f} (\alpha + \beta)\,(t_f - t_o)$$

and so $||Rx - Ry|| < q||x - y||$, $q < 1$, provided

$$q = (\alpha + \beta)\,(t_f - t_o) < 1$$

ie $(t_f - t_o) < \dfrac{1}{\alpha + \beta}$.

□

Comments

1. The matrix norms $|C(t, \tau)|$ and $|D(t,\tau)|$ can be computed as spectral norms or Euclidean norms. Recalling that the Euclidean length of $z(t)$ is given by

$$\|z(t)\| = (|z_1|^2 + |z_2|^2 + \ldots + |z_n|^2)^{\frac{1}{2}} \tag{71}$$

we define in the standard way (see, e.g., Wilkinson [62] p 56-57)

the spectral norm

$$\|C(t,t_o,\tau)\|_2 = (\text{max eigenvalue of } (C(t,t_o,\tau)^T C(t,t_o t))^{\frac{1}{2}} \tag{72}$$

and the Schur, or Euclidean norm

$$\|C(t,t_o,\tau)\|_E = (\sum_{i,j} |c_{ij}(t,t_o,\tau)|^2)^{\frac{1}{2}} \tag{73}$$

where it is well-known that

$$\|C\|_2 \leq \|C\|_E \leq n^{\frac{1}{2}} \|C\|_2. \tag{74}$$

2. Although it would be desirable to characterize further the time interval over which the general successive approximations scheme converges, this does not seem to be fruitful.

Recall that $C(t, \tau, t_o)$ is defined by (67) as

$$C(t, \tau) = \begin{bmatrix} \phi_{12}(t,t_o) \\ \phi_{22}(t,t_o) \end{bmatrix} \phi_{22}(t_f,t_o)^{-1} [\phi_{21}(t_{f_1},\tau) \; \phi_{22}(t_{f_1},\tau)] E(\tau)$$

$$= \begin{bmatrix} \phi_{12}(t,t_o)\phi_{22}(t_f,t_o)^{-1}\phi_{21}(t_{f_1},\tau) & \phi_{12}(t,t_o)\phi_{22}(t_f,t_o)^{-1}\phi_{22}(t_{f_1},\tau) \\ \phi_{22}(t,t_o)\phi_{22}(t_{f_1},t_o)^{-1}\phi_{21}(t_{f_1},\tau) & \phi_{22}(t,t_o)\phi_{22}(t_{f_1},t_o)^{-1}\phi_{22}(t_{f_1},\tau) \end{bmatrix} E(\tau)$$

... (75)

which we designate, for simplicity, $C = \phi E$.

Furthermore $D(t,\tau)$ is defined by (68) as

$$D(t,\tau) = \Phi(t,\tau)E(\tau)$$

$$= \begin{Bmatrix} \phi_{11}(t,\tau) & \phi_{12}(t,\tau) \\ \phi_{21}(t,\tau) & \phi_{22}(t,\tau) \end{Bmatrix} E(\tau) \qquad (76)$$

which we designate, for simplicity, $D = \Phi E$.

Hence, returning to the condition (70)

$$t_f - t_o < \frac{1}{\alpha+\beta} = \frac{1}{\sup_{t,\tau} |C(t,\tau)| + \sup_{t,\tau} |D(t,\tau)|} \quad ;$$

we have that

$$t_f - t_o < \frac{1}{\sup_{t,\tau} |\phi E| + \sup_{t,\tau} |\Phi E|}$$

All we can say is that, the smaller is $(\sup_{t,\tau} |\phi E| + \sup_{t,\tau} |\phi E|)$, the longer is the terminal time over which successive approximations will converge. This does not translate into any easy formula for choosing K.

3. This result says nothing about the *rate* of convergence, which might of course be very slow for some problems;

4. Since the proof uses only the linearity of $\Phi(t,\tau)$, the result extends without difficulty to time-varying systems A(t), B(t) etc.

Applicability and numerical computation of successive approximations scheme

1. Applicability

This follows immediately from the algorithm construction.

For if $\omega^k = \omega^{k+1} = \hat{\omega}$,

then $\dot{\hat{\omega}} = K\hat{\omega} + E\hat{\omega} = C\hat{\omega}$

i.e $\hat{\omega}$ satisfies the optimality conditions.

2. Numerical computation

We now consider how general 2PBVP (30) might be solved in practice. In general we employ the relation

$$\lambda(t) = P(t)\, x(t) + \phi(t) \tag{77}$$

and obtain a Riccati equation and a linear differential "tracking equation".

We have that

$$\begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix}^{k+1} = K \begin{pmatrix} x \\ \lambda \end{pmatrix}^{k+1} + E \begin{pmatrix} x \\ \lambda \end{pmatrix}^{k}, \qquad \begin{matrix} x^{k+1}(t_o) = x_o \\ \lambda^{k+1}(t_f) = 0 \end{matrix}$$

Setting $\lambda^{k+1} = Px^{k+1} + \phi^{k+1}$ for all k (77a)

we have that

$$\dot{\lambda}^{k+1} = P\dot{x}^{k+1} + \dot{P}x^{k+1} + \dot{\phi}^{k+1}$$

Then putting

$$\begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} = K, \quad \begin{pmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{pmatrix} = E$$

we have that

$$0 = -k_{21}x^{k+1} - k_{22}\lambda^{k+1} - e_{21}x^{k} - e_{22}\lambda^{k} + P[k_{11}x^{k+1} + k_{12}\lambda^{k+1} + e_{11}x^{k} + e_{12}\lambda^{k}] + \dot{P}x^{k+1} + \dot{\phi}^{k+1}$$

i.e.
$$0 = -k_{21}x^{k+1} - k_{22}Px^{k+1} - k_{22}\phi^{k+1} - e_{21}x^{k} - e_{22}Px^{k} - e_{22}\phi^{k} + Pk_{11}x^{k+1} + Pk_{12}Px^{k+1} + Pk_{12}\phi^{k+1} + Pe_{11}x^{k} + Pe_{12}Px^{k} + Pe_{12}\phi^{k} + \dot{P}x^{k+1} + \dot{\phi}^{k+1}$$

This is satisfied if

$$-\dot{P} = Pk_{11} - k_{22}P + Pk_{12}P - k_{21}, \quad P(t_f) = 0 \tag{78}$$

and
$$-\dot{\phi}^{k+1} = (-k_{22} + Pk_{12})\phi^{k+1} + (-e_{22}P + Pe_{11} + Pe_{12}P - e_{21})x^{k} + (Pe_{12} - e_{22})\phi^{k}, \quad \phi(t_f) = 0, \tag{79}$$

where we use the fact that P given by (77a) is independent of k.

Let us translate (78) and (79) into the Takahara case 1.

Substituting

$$K = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix}, \quad E = \begin{bmatrix} H & -GR^{-1}(B+G)^T - BR^{-1}G^T \\ 0 & -H^T \end{bmatrix}$$

we have that

$$-\dot{P} = PA - A^TP - PBR^{-1}B^TP + Q, \qquad P(t_f) = 0 \tag{80}$$

$$-\dot{\phi}^{k+1} = (A^T - PBR^{-1}B^T)\phi^{k+1} + (PH + H^TP - P[GR^{-1}(B+G)^T + BR^{-1}G^T]P)x^k + [H^T - P[GR^{-1}(B+G)^T) + BR^{-1}G^T)]\phi^k, \quad \phi(t_f) = 0 . \tag{81}$$

Finally let us translate (78) and (79) into the special case of the diagonal K matrix, given by (31) :

$$K = \begin{bmatrix} A & 0 \\ 0 & -A^T \end{bmatrix}, \qquad E = \begin{bmatrix} H & -(B+G)R^{-1}(B+G)^T \\ -Q & -H^T \end{bmatrix}$$

In this case we can solve the 2PBVP through two independent equations in x and λ, and do not need the relation (77) linking x and λ :

$$\lambda = Px + \phi .$$

This is the attractive feature of this scheme.

Note that to start the algorithm we choose an arbitrary $x^0(t)$ with $x^0(t_0) = x_0$, and an arbitrary $\lambda^0(t)$ with $\lambda^0(t_f) = 0$. Then ϕ^0 is given by

$$\lambda^0 = Px^0 + \phi^0$$

and an expression analogous to (79) can easily be derived.

CHAPTER 5

The Dynamic Interconnected Linear Quadratic Regulator : Operator Formulation

§1 Introduction and outline of chapter

§2 Operator formalism in Hilbert space setting

§3 Gradient-based algorithms

§4 Interaction prediction scheme

§5 Successive approximations and multilevel schemes

§1. Introduction

In this chapter the interconnected LQ regulator is recast from the standard differential description given in Chapter 4 to the language of operators acting on a Hilbert control space. The basic model is derived in §2, together with cost and gradient formulae that will be used repeatedly in the sequel. In §3 a range of gradient-based optimization algorithms is discussed with a view to their generating decomposition schemes. As in the case of the static regulator, such a systematic approach to decomposition has not previously been published. It turns out that a successful scheme is derived from the Newton-Raphson (K^{-1})/Contraction Mapping (K^{-1}) algorithm, as well as from the Conjugate Gradient (K^{-1}) algorithm. In both cases the notion of a "simple model" is exploited to generate a decomposition procedure.

In §4 an interaction prediction algorithm using interaction modification is derived, based largely on the work of Cohen [6]. It is then shown *directly* by looking at the fundamental control sequences generated by each algorithm, that this scheme is in fact identical with the NR(K^{-1}) scheme derived in §3, where K is chosen naturally.

Finally in §5 a wholly original operator version of the Takahara scheme is derived, giving yet another set of convergence conditions. The scheme is successfully generalized as Jacobi iteration applied to the operator system which constitutes the necessary conditions of optimality.

§2. Operator formalism in Hilbert space setting

We begin with the standard Linear Quadratic interconnected system

$$\dot{x}_i = A_i x_i + B_i u_i + z_i, \quad x_i(t_o) = x_{io} \tag{1}$$

$$z_i = \sum_{j=1}^{N} G_{ij} u_j + \sum_{j=1}^{N} H_{ij} x_j \tag{2}$$

with performance

$$G_i(u_i, x_i) = \tfrac{1}{2}\int_{t_o}^{t_f} (x_i^T Q_i x_i + u_i^T R_i u_i)\,dt \,. \tag{3}$$

In aggregate terms the system may be written as

$$\dot{x} = (A+H)x + (B+G)u, \quad x(t_o) = x_o \tag{4}$$

or in 'decoupled' form as

$$\dot{x} = Ax + Bu + z, \quad x(t_o) = x_o \tag{5}$$

$$z = Gu + Hx \tag{6}$$

with performance

$$G(u,x) = \tfrac{1}{2}\int_{t_o}^{t_f} (x^T Q x + u^T R u)\,dt. \tag{7}$$

Using the variation of constants formula we have , from (4),

$$x(t) = \Phi(t,t_o)x_o + \int_{t_o}^{t} \Phi(t,\tau)(B+G)u(\tau)\,d\tau \tag{8}$$

or

$$x(t) = x_s(t) + \int_{t_o}^{t} W(t,\tau)u(\tau)\,d\tau \,, \tag{9}$$

where

$$x_s(t) = \Phi(t,t_o)x_o \tag{10}$$

and

$$W(t,\tau) = \Phi(t,\tau)(B+G) \tag{11}$$

is the convolution map, and the transition map is defined by

$$\frac{d}{dt}\Phi(t,\tau) = (A+H)\Phi(t,\tau), \quad \Phi(t_o,t_o) = I. \tag{12}$$

Similarly from (5) we have

$$x(t) = \hat{\Phi}(t,t_o)x_o + \int_{t_o}^{t} \hat{\Phi}(t,\tau)(Bu(\tau) + z(\tau))\,d\tau \tag{13}$$

or

$$x(t) = \hat{x}_s(t) + \int_{t_o}^{t} \hat{W}(t,\tau)u(\tau)\,d\tau + \int_{t_o}^{t} \hat{\Phi}(t,\tau)z(\tau)\,d\tau, \tag{14}$$

where $\hat{x}_s(t) = \hat{\Phi}(t,t_o)x_o$,

and $\hat{W}(t,\tau) = \hat{\Phi}(t,\tau)B$ (15)

is the convolution map and the 'decoupled' transition map is defined by

$$\frac{d}{dt}\hat{\Phi}(t,\tau) = A\,\hat{\Phi}(t,\tau), \quad \hat{\Phi}(t_o,t_o) = I. \tag{16}$$

We now consider the functions $x(t)$, $u(t)$, $z(t)$ as elements of the Hilbert space $L_2[t_o,t_f]$, with inner product

$$\langle v,w\rangle \triangleq \int_{t_o}^{t_f} v^T(t)w(t)dt. \tag{17}$$

We define the operators Q, R, B, G, W, $\hat{W}$, Φ, $\hat{\Phi}$ by

$$\begin{aligned}
(Qv)(t) &= Qv(t) \quad \forall t\\
(Rx)(t) &= Rx(t) \quad \forall t\\
(Bv)(t) &= Bv(t) \quad \forall t\\
(Gv)(t) &= Gv(t) \quad \forall t\\
(Wv)(t) &= \int_{t_o}^{t} W(t,\tau)v(\tau)d\tau\\
(\hat{W}v)(t) &= \int_{t_o}^{t} \hat{W}(t,\tau)v(\tau)d\tau\\
(\Phi v)(t) &= \int_{t_o}^{t} \Phi(t,\tau)v(\tau)d\tau\\
(\hat{\Phi}v)(t) &= \int_{t_o}^{t} \hat{\Phi}(t,\tau)v(\tau)d\tau.
\end{aligned} \tag{18}$$

Note that we should differentiate between the operator W and its Kernel $W(t,\tau)$, etc, but in the sequel there will be no possibility of confusion so we do not use different symbols.(1)

Because of symmetry assumptions on matrices Q, R, the operators Q, R are self-adjoint, while it is readily shown that

(1) For the same reason we use transpose notation B^T, G^T etc to denote the adjoints B^*, G^* etc.

$$(W*v)(t) = \int_t^{t_f} W^T(\tau,t)v(\tau)d\tau$$

$$(\hat{W}*v)(t) = \int_t^{t_f} \hat{W}^T(\tau,t)v(\tau)d\tau \quad \text{etc} \tag{19}$$

i.e. W, $\hat{W}$, Φ, $\hat{\Phi}$ are not self-adjoint.

We now have the Hilbert space operator system

$$x = x_s + Wu \tag{20}$$

corresponding to (9), where

$$W = \Phi(B+G) \tag{21}$$

from (11), and the 'decoupled' system

$$x = \hat{x}_s + \hat{W}u + \hat{\Phi}z \tag{22}$$

corresponding to (14), where

$$\hat{W} = \hat{\Phi}B \tag{23}$$

from (15).

The criteria are, in aggregate form,

$$G(u,x) = \tfrac{1}{2}\langle x,Qx\rangle + \tfrac{1}{2}\langle u,Ru\rangle. \tag{24}$$

We note that in the Mesarovic notation employed in Chapter 2, (20) corresponds to

$$x = P(u)$$

while (22) corresponds to

$$x = \bar{P}(u,z),$$

while $z = Gu + Hx$ (25)

corresponds to $z = H(x,u)$.

Substituting (21) and (22) in (25) gives

$$z = Gu + H\Phi(B+G)u + Hx_s \tag{26}$$

which corresponds to $z = K(u)$.

The cost corresponding to the performance (24) is

$$J(u) = G(u,P(u)) = \tfrac{1}{2}\langle x_s, Qx_s\rangle + \tfrac{1}{2}\langle u,(R+W^*QW)u\rangle + \langle u, W^*Qx_s\rangle. \tag{27}$$

The system gradient is given by

$$g(u) = \frac{dJ}{du}(u) = (R+W^*QW)u + W^*Qx_s \tag{28}$$

while the second derivative, represented by $\frac{d^2J}{du^2}$, is given by

$$\frac{d^2J}{du^2} = R + W^*QW.$$

Reformulating (28) we obtain a 'feedback' form involving the system state

$$g(u) = Ru + W^*Qx^u \tag{29}$$

where $$x^u = Wu + x_s \tag{30}$$

or $$\dot{x}^u = (A+H)x^u + (B+G)u, \quad x(t_o) = x_o.$$

This can be further transformed to give a feedback form involving the system costate. Since from (21)

$$W = \Phi(B+G)$$

we have from (29) that

$$g(u) = Ru + (B+G)^T\lambda^u \tag{31}$$

where $$\lambda^u \triangleq \Phi^*Qx^u \tag{32}$$

or $$\dot{\lambda}^u = -(A+H)^T\lambda^u - Qx, \quad \lambda(t_f) = 0.$$

Alternatively we may derive a 'decoupled' or decentralized version of the system cost, which will yield decoupled formulae for the gradient. These formulae will be used repeatedly in the sequel:

Setting $\bar{J}(u,z) = G(u,\bar{P}(u,z))$ we have

$$\bar{J}(u,z) = \tfrac{1}{2}\langle \hat{x}_s + \hat{W}u + \hat{\Phi}z,\ Q(\hat{x}_s + \hat{W}u + \hat{\Phi}z)\rangle + \tfrac{1}{2}\langle u, Ru\rangle$$

so that

$$\bar{J}(u,z) = \tfrac{1}{2}\langle u,(R+\hat{W}^*Q\hat{W})u\rangle + \langle u,\hat{W}^*Q\hat{\Phi}z\rangle + \tfrac{1}{2}\langle z,\hat{\Phi}^*Q\hat{\Phi}z\rangle \tag{33}$$

$$+ \langle u,\hat{W}^*Q\hat{x}_s\rangle + \langle z,\hat{\Phi}^*Q\hat{x}_s\rangle + \tfrac{1}{2}\langle \hat{x}_s, Q\hat{x}_s\rangle$$

We may now compute partial gradients of the 'decoupled' cost (33). First we have that

$$\hat{g}(u,z) = \frac{\partial \bar{J}}{\partial u}(u,z) = (R+\hat{W}^*Q\hat{W})u + \hat{W}^*Q\hat{\Phi}z + \hat{W}^*Q\hat{x}_s \tag{34}$$

while the second partial derivative, represented by $\frac{\partial^2 \bar{J}}{\partial u^2}$, is given by

$$\frac{\partial^2 \bar{J}}{\partial u^2} = R + \hat{W}^*Q\hat{W}. \tag{35}$$

Reformulating (34) we obtain a 'feedback' form

$$\frac{\partial \bar{J}}{\partial u}(u) = Ru + \hat{W}^*Q\hat{x} \tag{36}$$

where
$$\hat{x} = \hat{W}u + \hat{\Phi}z + \hat{x}_s \tag{37}$$

or
$$\dot{\hat{x}} = A\hat{x} + Bu + z, \quad \hat{x}(t_o) = x_o.$$

Furthermore, from (15) we have that

$$\hat{W} = \Phi B$$

so from (36) we have that

$$\frac{\partial \bar{J}}{\partial u}(u) = Ru + B^T\hat{\lambda} \tag{38}$$

where we define the 'decoupled' costate by

$$\hat{\lambda} \triangleq \hat{\Phi}^*Q\hat{x} \tag{39}$$

or
$$\dot{\hat{\lambda}} = -A^T\hat{\lambda} - Qx, \quad \hat{\lambda}(t_f) = 0 \tag{40}$$

Finally we note the connection between the system gradient and Hamiltonian formulation. From the Hamiltonian

$$H(x,u,\lambda,t) = \tfrac{1}{2}\langle x,Qx\rangle + \tfrac{1}{2}\langle u,Ru\rangle + \langle \lambda,(A+H)x + (B+G)u\rangle \tag{41}$$

we have that

$$\frac{\partial H}{\partial u}(t) = Ru(t) + (B+G)^T\lambda(t)$$

where

$$\dot{\lambda}(t) = -(A+H)^T\lambda(t) - Qx(t), \quad \lambda(t_f) = 0 \tag{43}$$

and from (31), we note that

$$\frac{dJ}{du} = \frac{\partial H}{\partial u} .$$

Similarly for the 'interconnected' Hamiltonian

$$\hat{H}(x,u,\lambda,z,t) = \tfrac{1}{2}\langle x,Qx\rangle + \tfrac{1}{2}\langle u,Ru\rangle + \langle\hat{\lambda},\ Ax + Bu + z\rangle \tag{44}$$

we have that

$$\frac{\partial\hat{H}}{\partial u} = Ru(t) + B^{T}\hat{\lambda}(t) \tag{45}$$

where

$$\dot{\hat{\lambda}}(t) = -A^{T}\hat{\lambda}(t) - Qx(t), \quad \hat{\lambda}(t_f) = 0$$

and from (37), we note that

$$\frac{\partial\bar{J}}{\partial u} = \frac{\partial\hat{H}}{\partial u} .$$

§3. Gradient algorithms

Just as we did in the static case, we may now use the Hilbert space model with cost (24); system (20) and decomposed system (22)derived in §2 to formulate function space gradient algorithms which generate decomposition schemes. We consider the Contraction Mapping (K^{-1}) / Newton Raphson (K^{-1}) scheme first.

(i) $CM(K^{-1})/NR(K^{-1})$ algorithm

Recall from Chapter 3.5 that in the case of quadratic costs the $CM(K^{-1})$ and $NR(K^{-1})$ algorithms are indentical. In this dynamic case the general cost

$$J(u) = a + \langle b,u\rangle + \tfrac{1}{2}\langle u,Cu)\rangle \tag{46}$$

is given by the particular cost (27)

$$J(u) = \tfrac{1}{2}\langle x_s,Qx_s\rangle + \langle W^{*}Qx_s,u\rangle + \tfrac{1}{2}\langle u,(R+W^{*}QW)u\rangle \tag{47}$$

so that in (46),

$$a = \tfrac{1}{2}\langle x_s,Qx_s\rangle$$

$$b = W^{*}Qx_s$$

$$C = R+W^{*}QW.$$

Let us approximate c by K, where

$$K = R + \hat{W}^* Q \hat{W}, \tag{48}$$

the motivation being that a suitable K will generate a decomposition scheme.

Using the $NR(K^{-1})$ notation, the update step is

$$u^{k+1} = u^k - K^{-1} g(u^k) \tag{49}$$

or $$u^{k+1} = u^k - (R + \hat{W}^* Q \hat{W})^{-1} g(u^k) \ .$$

Whereas in the static case it was not an insuperable problem to algebraically invert $(R + \hat{W}^T Q \hat{W})$, in the dynamic case the operator inversion of $(R + \hat{W}^* Q \hat{W})$ would be a formidable task. Hence Allwright's notion of 'simple model' comes into its own here. We need to evaluate

$$K^{-1} g = h,$$

which is equivalent to solving

$$Kh = g,$$

i.e. $$(R + \hat{W}^* Q \hat{W}) h = g \ .$$

This is the optimality condition for the optimal control problem

$$\left.\begin{aligned} &\min_h \tfrac{1}{2} \langle h, (R + \hat{W}^* Q \hat{W}) h \rangle - \langle g, h \rangle \\ \text{i.e.} \quad &\min_h \tfrac{1}{2} \langle x, Qx \rangle + \tfrac{1}{2} \langle h, Rh \rangle - \langle g, h \rangle \\ &\text{s.t.} \quad x = \hat{W} h \\ &\text{or} \quad \dot{x} = Ax + Bh, \quad x(t_o) = 0 \ . \end{aligned}\right\} \tag{50}$$

In Allwright's terminology, (50) is a 'simple model' of the original problem, viz

$$\left.\begin{aligned} &\min_u \tfrac{1}{2} \langle x, Qx \rangle + \tfrac{1}{2} \langle u, Ru \rangle . \\ &\text{s.t.} \quad x = Wu + x_s \\ &\text{or} \quad \dot{x} = (A+H)x + (B+G)u, \quad x(t_o) = x_o . \end{aligned}\right\} \tag{51}$$

Note how (51) has been modified in relation to (50); this is a point we shall return to in the context of interaction operators. Meanwhile we solve (50).

This is a slightly non-standard LQ problem. We solve it through use of a 'diagonal' Riccati equation and a linear differential 'tracking' equation, as follows.

Writing the Hamiltonian for (50) as

$$H = \tfrac{1}{2}\langle x,Qx\rangle + \tfrac{1}{2}\langle h,Rh\rangle - \langle g,h\rangle + \langle p,Ax+Bh\rangle \tag{52}$$

we have, by Maximum Principle, the necessary conditions for optimality:

$$\begin{aligned} \frac{\partial H}{\partial p} &= \dot{x} = Ax + Bh \\ \frac{\partial H}{\partial x} &= -\dot{p} = Qx + A^T p \\ \frac{\partial H}{\partial u} &= 0 = Rh - g + B^T p \end{aligned} \tag{53}$$

Hence $\hat{h} = -R^{-1}B^T p + R^{-1}g$

so that

$$\begin{aligned} \dot{x} &= Ax - BR^{-1}B^T p + BR^{-1}g, \qquad x(t_o) = x_o \\ \dot{p} &= -Qx - A^T p \qquad\qquad , \qquad p(t_f) = 0 \end{aligned} \tag{54}$$

We end up then with a non-standard 2PBVP (54). This can be exploited to give a 'feedback' solution, by setting

$$p = Px + \phi \tag{55}$$

so that

$$\dot{p} = P\dot{x} + \dot{P}x + \dot{\phi} \tag{56}$$

Substituting (54) into (56) gives

$$0 = Qx + A^T Px + A^T\phi + PAx - PBR^{-1}B^T Px - PBR^{-1}B^T\phi + PBR^{-1}g + \dot{P}x + \dot{\phi} \quad .$$

Since this is true for all x, we have that

$$-\dot{P} = PA + A^T P - PBR^{-1}B^T P + Q, \qquad P(t_f) = 0 \tag{57}$$

$$-\dot{\phi} = (A^T - PBR^{-1}B^T)\phi + PBR^{-1}g. \qquad \phi(t_f) = 0 \tag{58}$$

Hence given g, the current gradient value, the solution to (50), and hence the value of $K^{-1}g$, is given by the equations (57), (58), (54), (55), and (53). The important point to note is that each of these expressions can be decomposed, so the full solution for the i^{th} subsystem is given by :

(i) Evaluate $P_i(t)$ once and for all, through the Riccati equation (57)

$$-\dot{P}_i = P_i A_i + A_i^T P_i - P_i B_i R_i^{-1} B_i^T P_i + Q_i, \; P_i(t_f) = 0.$$

(ii) Evaluate $\phi_i^k(t)$ through the 'tracking equation' (58) :

$$-\dot{\phi}_i^k = (A_i^T - P_i B_i R_i^{-1} B_i^T)\phi_i^k + P_i B_i R_i^{-1} g_i^k, \quad \phi_i(t_f) = 0$$

(iii) Evaluate $x_i^k(t)$ through the state equation (54) :

$$\dot{x}_i = (A_i - B_i R_i^{-1} B_i^T P_i) x_i^k - B_i R_i^{-1} B_i^T \phi_i^k + B_i R_i^{-1} g_i^k \quad , \quad x_i^k(t_o) = 0$$

(iv) Evaluate $p_i^k(t)$ through the linear relation (55)

$$p_i^k(t) = P_i(t) x_i^k(t) + \phi_i^k(t)$$

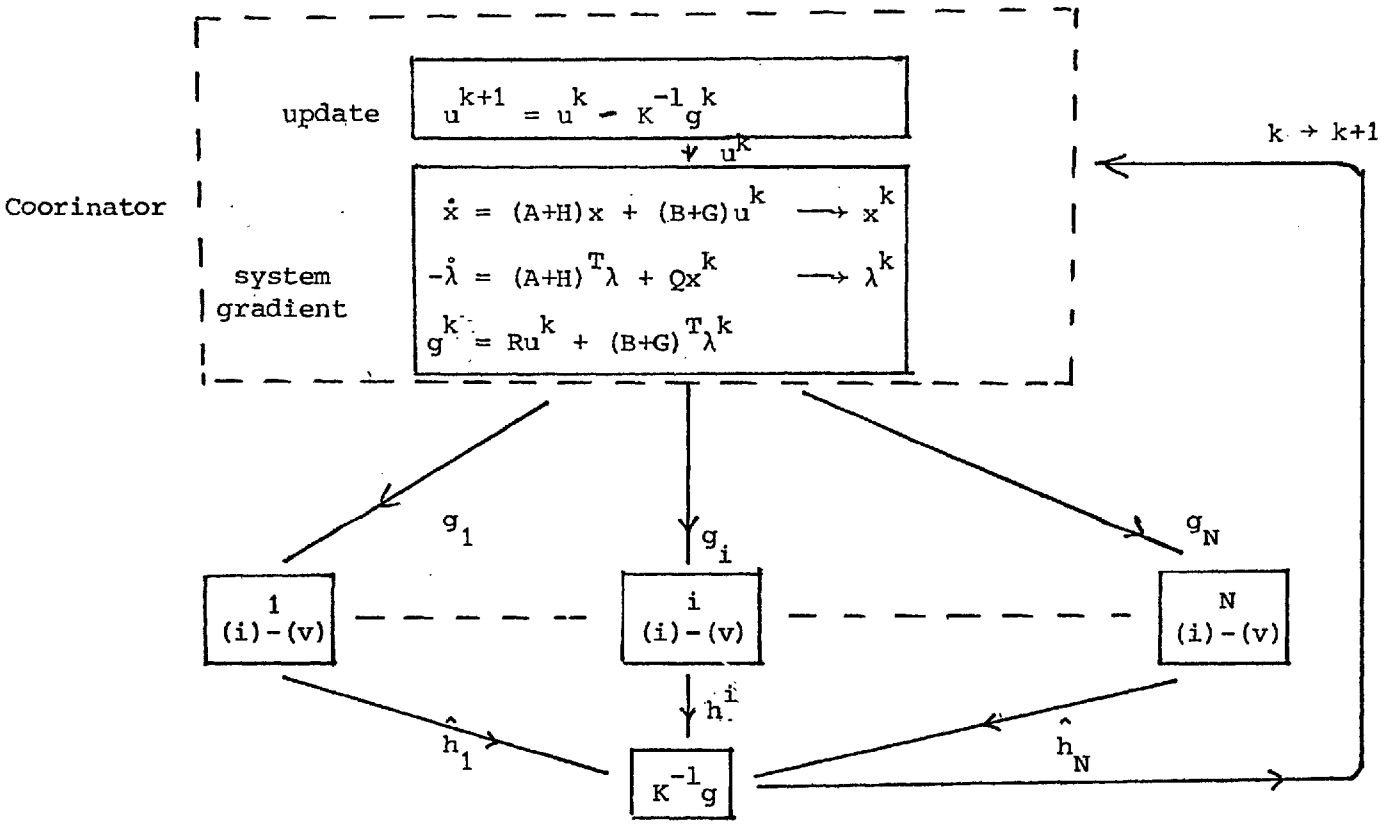
(v) Evaluate $h_i^k(t)$ through the stationarity condition (53)

$$h_i^k(t) = -R_i^{-1} B_i^T p_i^k(t) + R_i^{-1} g_i^k(t)$$

and finally

(vi) Set $K^{-1}g^k = h^k$.

Hence the flow of information in the decomposed NR(K^{-1}) scheme is as follows.



Remark

The coordinator in this scheme has the heavy task of solving the state equation forwards, and costate equation backwards, to compute the system gradient. This is exactly the same computational load as envisaged by Cohen [6] in the 'pure reinjection' algorithm considered below. Once the system gradient has been computed, the subsystems compute their portion of $K^{-1}g^k$, and the coordinator updates the control.

We now examine the convergence of the above schemes. Recall that a sufficient condition for convergence, based on the identity between $NR(K^{-1})$ and $CM(K^{-1})$, is

$$\|K^{-1}E\| < 1.$$

Now in this case

$$K = R + \hat{W}^*Q\hat{W}$$

so $$E = C - K = W^*QW - \hat{W}^*Q\hat{W}$$

and the convergence condition is :

$$\|(R+\hat{W}^*Q\hat{W})^{-1}(W^*QW-\hat{W}^*Q\hat{W})\| < 1. \tag{59}$$

Alternatively we require that

$$\|K^{-1}C - I\| < 1$$

which is equivalent to

$$K - \tfrac{1}{2}C > 0. \tag{60}$$

We recall that (60) is the coercivity condition given by Cohen [7] for his algorithm.

It is also interesting to note, after Cohen, that the cost $J(u^k)$ decreases strictly with u^k with convergence rate

$$J(u^k) - J(u^{k+1}) = \langle u^k-u^*, z(u^k-u^*)\rangle$$

where $$z = CK^{-1}(K-\tfrac{1}{2}C)K^{-1}C.$$

It is easy to see why this is so, as checked in Chapter 3. We have that

$$J(u) = a + \langle b,u\rangle + \tfrac{1}{2}\langle u,Cu\rangle.$$

Also $$\frac{dJ}{du}(u^k) = Cu^k + b$$

and $$\frac{dJ}{du}(u^*) = Cu^* + b = 0$$

so that $$\frac{dJ}{du}(u^k) = C(u^k-u^*).$$

Then the NR(K^{-1}) update formula is

$$u^{k+1} - u^k = K^{-1}C(u^k-u^*).$$

Expanding $J(u^{k+1})$ to 2nd order about u^k we have

$$J(u^{k+1}) = J(u^k) + \langle \nabla J(u^k), u^{k+1}-u^k\rangle + \tfrac{1}{2}\langle u^{k+1}-u^k, \frac{d^2J}{du^2}(u^{k+1}-u^k)\rangle$$

i.e. $$J(u^{k+1})-J(u^k) = \langle C(u^k-u^*), -K^{-1}C(u^k-u^*)\rangle + \tfrac{1}{2}\langle K^{-1}C(u^k-u^*), CK^{-1}C(u^k-u^*)\rangle$$

$$= \langle u^k-u^*, (\tfrac{1}{2}CK^{-1}CK^{-1}C - CK^{-1}C)(u^k-u^*)\rangle$$

$$= -\langle u^k-u^*, (CK^{-1}(K-\tfrac{1}{2}C)K^{-1}C)(u^k-u^*)\rangle$$

and again we confirm that

$$J(u^{k+1}) - J(u^k) < 0 \text{ provided } C > 0$$

$$K > 0$$

$$\text{and} \quad K - \tfrac{1}{2}C > 0.$$

Note that we ensure $C > 0$, $K > 0$ by the original convexity assumptions : $Q \geq 0$, $R > 0$. Thus we have obtained a sound decomposition scheme for solving the overall system, which converges to the optimal control if the coercivity condition (60) is satisfied.

(ii) $CG(K^{-1})$ algorithm

Recall from Chapter 3.5 that this algorithm consists of the basic steps :

0. $s^o = -K^{-1}g^o, \quad k = 0$

1. $u^{k+1} = u^k + ws^k$, where w is the optimal step length

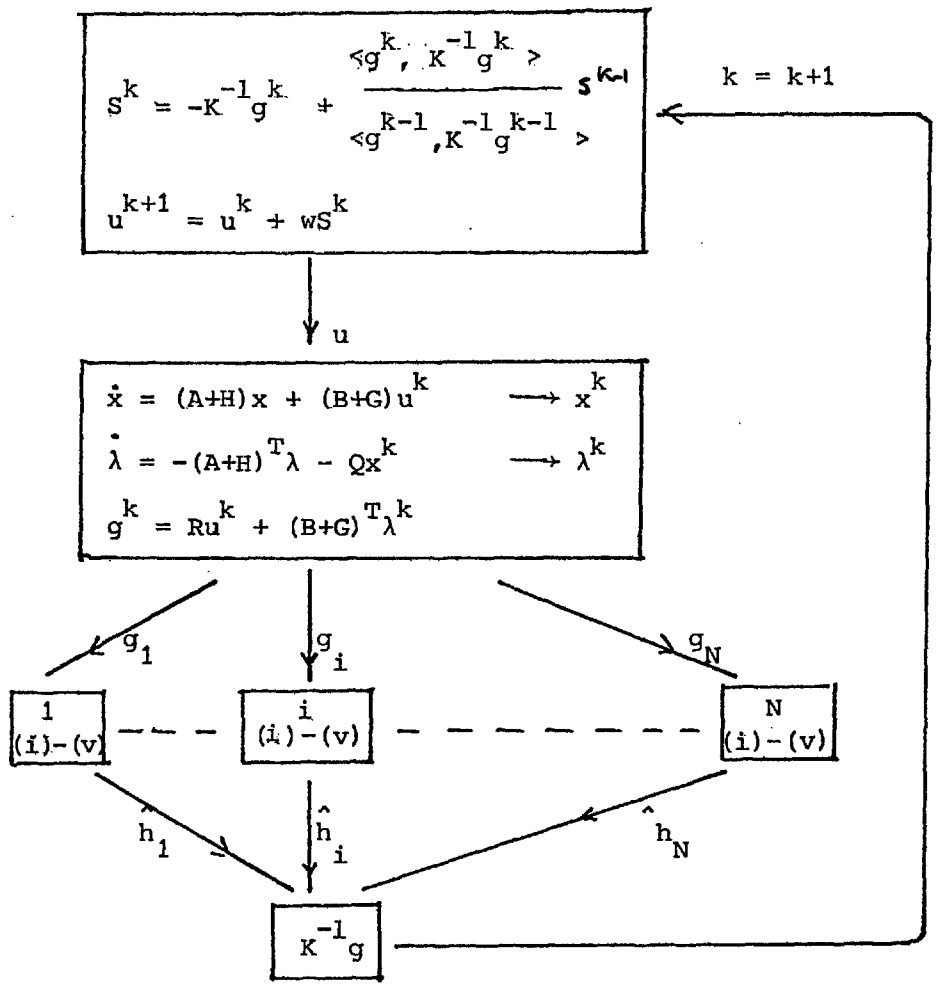
2. $s^{k+1} = -K^{-1}g^{k+1} + \beta s^k$

$$\beta = \frac{\langle g^{k+1}, K^{-1}g^{k+1}\rangle}{\langle g^k, K^{-1}g^k\rangle}$$

3. $k = k+1$

Again we set $K = R + \hat{W}^*Q\hat{W}$. Then we have the remarkable fact that the update expression requires the computation of $K^{-1}g$, where where K is the same as in the $NR(K^{-1})$ algorithm. Hence exactly the same infimal computations will be performed, but the coordinator will have a different update procedure. It was in the context of the function space $CG(K^{-1})$ algorithm that Allwright [47] introduced the notion of finding $K^{-1}g$ through a 'simple model'

Hence we have the information flow



In [51] Allwright discusses convergence of this $CG(K^{-1})$ algorithm, and shows that it converges for all decompositions $K = K^T > 0$ and moreover it has a potential advantage over the $NR(K^{-1})/CM(K^{-1})$ algorithm in terms of speed of convergence. Hence despite the heavy computational load imposed on the coordinator, updating the system gradient at each step, this is a very attractive decomposition procedure.

Hence we have succeeded in deriving two decomposition schemes based on the system gradient. These were derived in the Hilbert space setting , but have been transcribed into a series of computations involving local state, costate and control trajectories.

§4. Interaction Prediction using Interaction Operator

We now derive an interaction prediction algorithm for the dynamic LQ regulator, following Cohen [6] and Cohen and Bernhard [7]. This scheme uses an interaction operator to modify, and thereby coordinate, subsystem criteria.

Recall from Chapter 2 that given a current control u^k, there is determined a current state $x^k = P(u^k)$ and costate λ^k and current interaction prediction $z^k = H(u^k, x^k)$. The interaction operator $\Gamma(u^k)$ is then used to modify local criteria, giving the full problem :

$$\begin{aligned} &\min_u G(x,u) + \Gamma(u^k)u \\ \text{s.t.}\quad &x = \bar{P}(u,z) \\ &z = K(u^k) \end{aligned} \tag{61}$$

where $$\Gamma(u^k) = \frac{\partial \bar{J}}{\partial z}(u^k, K(u^k)) * \frac{dK}{du}(u^k). \tag{62}$$

We specify a simple 'reinjection' procedure to update the control, whereby we take u^{k+1} as $\hat{u}(\alpha^k)$, the optimal control found at step k, for a given interaction prediction α^k.

The goal of this section will be to derive a concrete representation of the interaction operator, in terms of system variables, and then to use this representation to derive the decomposed solution to the modified system (61). From this we shall obtain the fundamental control update expression determined by this procedure, as well as a set of formulae suitable for numerical computation. We shall then show that this IP scheme is in fact identical to the Newton Raphson (K^{-1}) scheme obtained above, where $K = R + \hat{W}^* Q \hat{W}$. Finally the relations between the computational scheme employed by the IP algorithm, and that employed

by the 'simple model' approach sketched above, will be examined.

Recall our basic model is as follows :

$$x = x_s + \Phi(B+G)u = P(u) \tag{63}$$

$$x = \hat{x}_s + \hat{\Phi}Bu + \hat{\Phi}z = \bar{P}(u,z) \tag{64}$$

$$z = Gu + Hx \qquad = H(x,u) \tag{65}$$

$$\therefore \quad z = [G+H\Phi(B+G)]u + Hx_s = K(u) \tag{66}$$

The criteria are, in aggregate form,

$$G(u,x) = \tfrac{1}{2}\langle x,Qx\rangle + \tfrac{1}{2}\langle u,Ru\rangle$$

so
$$J(u) = G(u,P(u)) = \tfrac{1}{2}\langle x_s,Qx_s\rangle + \tfrac{1}{2}\langle u,[R+(B+G)^T\Phi^*Q\Phi(B+G)]u\rangle + \langle u,(B+G)^T\Phi^*Qx_s\rangle \tag{67}$$

while the 'decoupled' version is

$$\bar{J}(u,z) = G(u,\bar{P}(u,z)) = \tfrac{1}{2}\langle u,(R+B^T\hat{\Phi}^*Q\hat{\Phi}B)u\rangle + \tfrac{1}{2}\langle z,\hat{\Phi}^*Q\hat{\Phi}z\rangle + \langle u,B^T\hat{\Phi}^*Q\hat{\Phi}z\rangle + \langle u,B^T\hat{\Phi}^*Q\hat{x}_s\rangle + \langle z,\hat{\Phi}^*Q\hat{x}_s\rangle + \tfrac{1}{2}\langle\hat{x}_s,Q\hat{x}_s\rangle \quad .... \tag{68}$$

We proceed to the derivation of the row vector $\Gamma(u)$.

We have from (68) that

$$\frac{\partial\bar{J}}{\partial z}(u,z) = \hat{\Phi}^*Q\hat{\Phi}z + \hat{\Phi}^*Q\hat{\Phi}Bu + \hat{\Phi}^*Q\hat{x}_s. \tag{69}$$

Substituting for z from (26) we have that

$$\frac{\partial\bar{J}}{\partial z}(u,z) = \{\hat{\Phi}^*Q\hat{\Phi}B + \hat{\Phi}^*Q\hat{\Phi}[G+H\Phi(B+G)]\}u + (\hat{\Phi}^*Q\hat{\Phi}H + \hat{\Phi}^*Q)\hat{x}_s. \tag{70}$$

From (26) we have that

$$\frac{dK}{du} = G + H\Phi(B+G) \tag{71}$$

Hence substituting (70), (71) in (62) gives

$$\Gamma(u) = u^T\{B^T\hat{\Phi}^*Q\hat{\Phi} + [G^T + (B+G)^T\Phi^*H^T]\hat{\Phi}^*Q\hat{\Phi}\}(G + H\Phi(B+G) + x_s^T(H^T\hat{\Phi}^*Q\hat{\Phi} + Q\hat{\Phi})(G + H\Phi(B+G)). \tag{72}$$

Expression (72) is clumsy, so we proceed analogously to Cohen [6,7] and obtain a simpler expression for $\Gamma(u)$ in terms of the system state and costate variables x and λ.

We have the following convenient

Lemma (Cohen)

$$\Gamma(\bar{u})u = \langle\bar{\lambda}, Hx + Gu\rangle - \langle\bar{\lambda}, H\hat{x}_s + H\hat{\Phi}z\rangle \tag{73}$$

where $\bar{u}$ is a given control,

$$\bar{x} = x_s + \Phi(B+G)\bar{u} \tag{74}$$

$$\bar{\lambda} = \Phi^* Q\bar{x} \tag{75}$$

$$x = \hat{x}_s + \hat{\Phi}Bu + \hat{\Phi}z \tag{76}$$

Proof

$$\text{Now} \quad \left(\frac{\partial\bar{J}}{\partial z}\right)^T = \left(\frac{\partial G}{\partial x}\right)^T \frac{\partial\bar{P}}{\partial z} \tag{77}$$

and we have, from

$$G(u,x) = \tfrac{1}{2}\langle x, Qx\rangle + \tfrac{1}{2}\langle u, Ru\rangle$$

$$\text{that} \qquad \frac{\partial G}{\partial x} = Qx; \tag{78}$$

while from

$$\bar{P}(u,z) = \hat{x}_s + \hat{\Phi}Bu + \hat{\Phi}z \tag{79}$$

$$\text{that} \qquad \frac{\partial\bar{P}}{\partial z} = \hat{\Phi} \quad .$$

Hence

$$\left(\frac{\partial\bar{J}}{\partial z}\right)^T = x^T Q\hat{\Phi} \quad , \tag{80}$$

so from (62) ,

$$\Gamma(\bar{u})u = \bar{x}^T Q\hat{\Phi}[G+H\Phi(B+G)]u$$

$$\text{or} \quad \Gamma(\bar{u})u = \langle Q\bar{x}, [\hat{\Phi}G+\hat{\Phi}H\Phi(B+G)]u\rangle. \tag{81}$$

We now employ the identity

$$\hat{\Phi}H\Phi(B+G)u = (\Phi-\hat{\Phi})(B+G)u = \Phi H\hat{\Phi}(B+G)u \tag{82}$$

which is easily obtained from setting the overall state x, given by (74), equal to the decoupled state x given by (76), after substituting for z by

$$z = Gu + Hx$$

or $$z = Gu + H\Phi(B+G)u + Hx_s .$$

Indeed, in (81)

$$\hat{\Phi}Gu + \hat{\Phi}H\Phi(B+G)u$$

$$= \hat{\Phi}Gu + (\Phi-\hat{\Phi})(B+G)u \qquad \text{by (82)(i)}$$

$$= (\Phi-\hat{\Phi})Bu + \Phi Gu$$

$$= \Phi H\hat{\Phi}Bu + \Phi Gu \qquad \text{by (82)(ii)}$$

so we have that

$$\Gamma(\bar{u})u = \langle Q\bar{x}, \Phi[Gu+H\hat{\Phi}Bu]\rangle \tag{83}$$

and rearranging (83) we obtain

$$\Gamma(\bar{u})u = \langle \Phi^*Q\bar{x},\ H\hat{\Phi}Bu+Gu \rangle . \tag{84}$$

We recognize $\hat{\Phi}Bu$ as $x - \hat{x}_s - \hat{\Phi}z$, from (76).

We recognize $\Phi^*Q\bar{x}$ as the system costate $\bar{\lambda}$.

Hence the Lemma is proved. ‖

We may now fill in the details of the modified optimization problem (61)

$$\begin{aligned} &\min_u \tfrac{1}{2}\langle x, Qx\rangle + \tfrac{1}{2}\langle u, Ru\rangle + \langle \lambda^k, (Hx + Gu)\rangle \\ &\text{s.t.} \quad x = \hat{\Phi}Bu + \hat{\Phi}z + \hat{x}_s \\ &\qquad z = Gu^k + Hx^k \end{aligned} \tag{85}$$

where we delete the cost term $-\langle \lambda^k, H\hat{x}_s + H\hat{\Phi}z\rangle$ which is independent of u.

This decomposes into N interconnected problems

$$\begin{aligned} &\min_{u_i} \tfrac{1}{2}\langle x_i, Q_i x_i\rangle + \tfrac{1}{2}\langle u_i, R_i u_i\rangle + \langle \lambda^k, (\textstyle\sum H_{ij}x_j + \sum G_{ij}u_j)\rangle \\ &\text{s.t.} \quad x_i = \hat{\Phi}_i B_i u_i + \hat{\Phi}_i z_i + \hat{x}_{si} \\ &\qquad z_i = \textstyle\sum G_{ij}u_j^k + \sum H_{ij}x_j^k \end{aligned} \tag{86}$$

and we may transpose the cost function to give i^{th} term

$$\min_{u_i} \tfrac{1}{2}\langle x_i, Q_i x_i\rangle + \tfrac{1}{2}\langle u_i, R_i u_i\rangle + \langle x_i, \sum H_{ji}^T \lambda_j^k + u_i^T \sum G_{ji}^T \lambda_j^k\rangle \qquad (87)$$

We solve the aggregate problem (85). The operator Lagrangian is :

$$L(x,u,\lambda) = \tfrac{1}{2}\langle x, Qx\rangle + \tfrac{1}{2}\langle u, Ru\rangle + \langle \lambda^k, (Hx + Gu)\rangle + \langle \lambda, (x - \hat{\Phi}Bu - \hat{\Phi}Gu^k - \hat{\Phi}Hx^k - \hat{x}_s)\rangle \qquad (88)$$

and the necessary conditions for optimality are :

$$\frac{\partial L}{\partial \lambda} = 0 = x - \hat{\Phi}Bu - \hat{\Phi}Gu^k - \hat{\Phi}Hx^k - \hat{x}_s \qquad (89)$$

$$\frac{\partial L}{\partial x} = 0 = Qx + \lambda + H^T\lambda^k \qquad (90)$$

$$\frac{\partial L}{\partial u} = 0 = Ru - B^T\hat{\Phi}^*\lambda + G^T\lambda^k \qquad (91)$$

By (91)

$$u = R^{-1}B^T\hat{\Phi}^*\lambda - R^{-1}G^T\lambda^k \qquad (92)$$

and by (90)

$$\lambda = -Qx - H^T\lambda^k. \qquad (93)$$

Substituting λ in (92) from (93) gives

$$u = -R^{-1}B^T\hat{\Phi}^*Qx - R^{-1}B^T\hat{\Phi}^*H^T\lambda^k - R^{-1}G^T\lambda^k \qquad (94)$$

and substituting for x from (89) gives

$$u = -R^{-1}B^T\hat{\Phi}^*Q\hat{\Phi}Bu - R^{-1}B^T\Phi^*Q\hat{\Phi}Gu^k - R^{-1}B^T\hat{\Phi}^*Q\hat{\Phi}Hx^k - (R^{-1}B^T\hat{\Phi}^*H^T + R^{-1}G^T)\lambda^k - R^{-1}B^T\hat{\Phi}^*Q\hat{x}_s. \qquad (95)$$

Now we know that

$$\lambda^k = \Phi^*Q\Phi(B+G)u^k + \Phi^*Qx_s \qquad (96)$$

and $$x^k = \Phi(B+G)u^k + x_s \qquad (97)$$

Substituting these values in (95) gives the update expression

$$u^{k+1} = -[I + R^{-1}B^T\hat{\Phi}^*Q\hat{\Phi}B]^{-1}\{R^{-1}B^T\Phi^*Q\hat{\Phi}Gu^k + R^{-1}B^T\hat{\Phi}^*Q\hat{\Phi}H\Phi(B+G)u^k \ldots$$

$$.... + (R^{-1}B^T\hat{\Phi}^*H^T + R^{-1}G^T)\Phi^*Q\Phi(B+G)u^k\}$$
$$-[I + R^{-1}B^T\hat{\Phi}^*Q\hat{\Phi}B]^{-1}\{R^{-1}B^T\hat{\Phi}^*Q\hat{x}_s + R^{-1}B^T\hat{\Phi}^*Q\hat{\Phi}Hx_s\} \quad , \tag{98}$$

and removing the factor R^{-1} :

$$u^{k+1} = -[R + B^T\hat{\Phi}^*Q\hat{\Phi}B]^{-1}\{B^T\Phi^*Q\hat{\Phi}G + B^T\hat{\Phi}^*Q\hat{\Phi}H\Phi(B+G) + (B^T\hat{\Phi}^*H^T + G^T)\Phi^*Q\Phi(B+G)\}u^k$$
$$-[R + B^T\hat{\Phi}^*Q\hat{\Phi}B]^{-1}\{B^T\hat{\Phi}^*Q\hat{x}_s + B^T\hat{\Phi}^*Q\hat{\Phi}Hx_s\} \quad . \tag{99}$$

Equation (99) is the fundamental update expression defining the control sequence mapped out by the interation prediction method with interaction operator modification and 'reinjection coordination'.

Numerical computation

Let us consider how the method might actually be employed. Going back to the necessary conditions (89) - (91) we have, *in differential terms*,

$$\dot{x} = Ax + Bu + Gu^k + Hx^k, \quad x(t_o) = x_o \tag{100}$$
$$\dot{\lambda} = -A^T\lambda - Qx - H^T\lambda^k \quad , \quad \lambda(t_f) = 0 \tag{101}$$
$$u = -R^{-1}B^T\lambda - R^{-1}G^T\lambda^k \tag{102}$$

so that

$$\dot{x} = Ax - BR^{-1}B^T\lambda - BR^{-1}G^T\lambda^k + Gu^k + Hx^k \quad , \quad x(t_o) = x_o \tag{103}$$
$$\dot{\lambda} = -Qx - A^T\lambda - H^T\lambda^k \quad , \quad \lambda(t_f) = 0. \tag{104}$$

We solve the 2PBVP (103) - (104) by setting

$$\lambda = Px + \phi \tag{105}$$

and getting a Riccati equation in P and linear differential 'tracking' equation in ϕ.

$$\text{From (105)}, \ \dot{\lambda} = P\dot{x} + \dot{P}x + \dot{\phi} \tag{106}$$

so that substituting (101), (102) gives

$$0 = Qx + A^TPx + A^T\phi + H^T\lambda^k + PAx - PBR^{-1}B^TPx - PBR^{-1}B^T\phi$$
$$-PBR^{-1}G^T\lambda^k + PGu^k + PHx^k + \dot{P}x + \dot{\phi}$$

Since this is true for all x, we have :

$$-\dot{P} = PA + A^T P - PBR^{-1}B^T P + Q \,, \qquad P(t_f) = 0 \tag{107}$$

$$-\dot{\phi} = (A^T - PBR^{-1}B^T)\phi - PBR^{-1}G^T\lambda^k + P(Gu^k + Hx^k) - H^T\lambda^k \,, \quad \phi(t_f) = 0. \tag{108}$$

Hence, given the current values of control u^k, system state x^k and costate λ^k, the solution to the IP equations is given by (105-108). The important point to note, as in the formulation of the NR(K^{-1}) algorithm above, is that each of these expressions can be decomposed. Hence the full decomposed algorithm consists of the following steps (Cohen and Bernhard (7)), for i = 1...N :

1) Integrate the Riccati equations :

$$-\dot{P}_i = P_i A_i + A_i^T P_i - P_i B_i R_i^{-1} B_i^T P_i + Q_i \,, \quad P_i(t_f) = 0 \;; \tag{109}$$

2) Select an initial control u^o, and set k = 0 ;

3) Integrate the system

$$\dot{x} = (A+H)x + (B+G)u^k, \quad x(t_o) = x_o \tag{110}$$

or apply u^k to the real system and observe the state x^k ;

4) Integrate the adjoint system

$$\dot{\lambda} = -(A+H)^T\lambda - Qx, \qquad \lambda(t_f) = 0 \tag{111}$$

to obtain λ^k ;

5) Integrate the tracking equations

$$-\dot{\phi}_i = (A_i^T - P_i B_i R_i^{-1} B_i^T)\phi_i - P_i B_i R_i^{-1} \sum G_{ij}\lambda_j^k + P\sum G_{ij} h_j^k + P\sum H_{ij} x_j^k$$

$$-\sum H_{ji}^T \lambda_j^k \,, \qquad \phi_i(t_f) = 0 \;; \tag{112}$$

6) Simulate the local processes with the controls :

$$u_i = -R_i^{-1} B_i^T P_i x_i - R_i^{-1} B_i \phi_i - R_i^{-1} \sum G_{ji}\lambda_j^k \tag{113}$$

i.e. integrate the systems

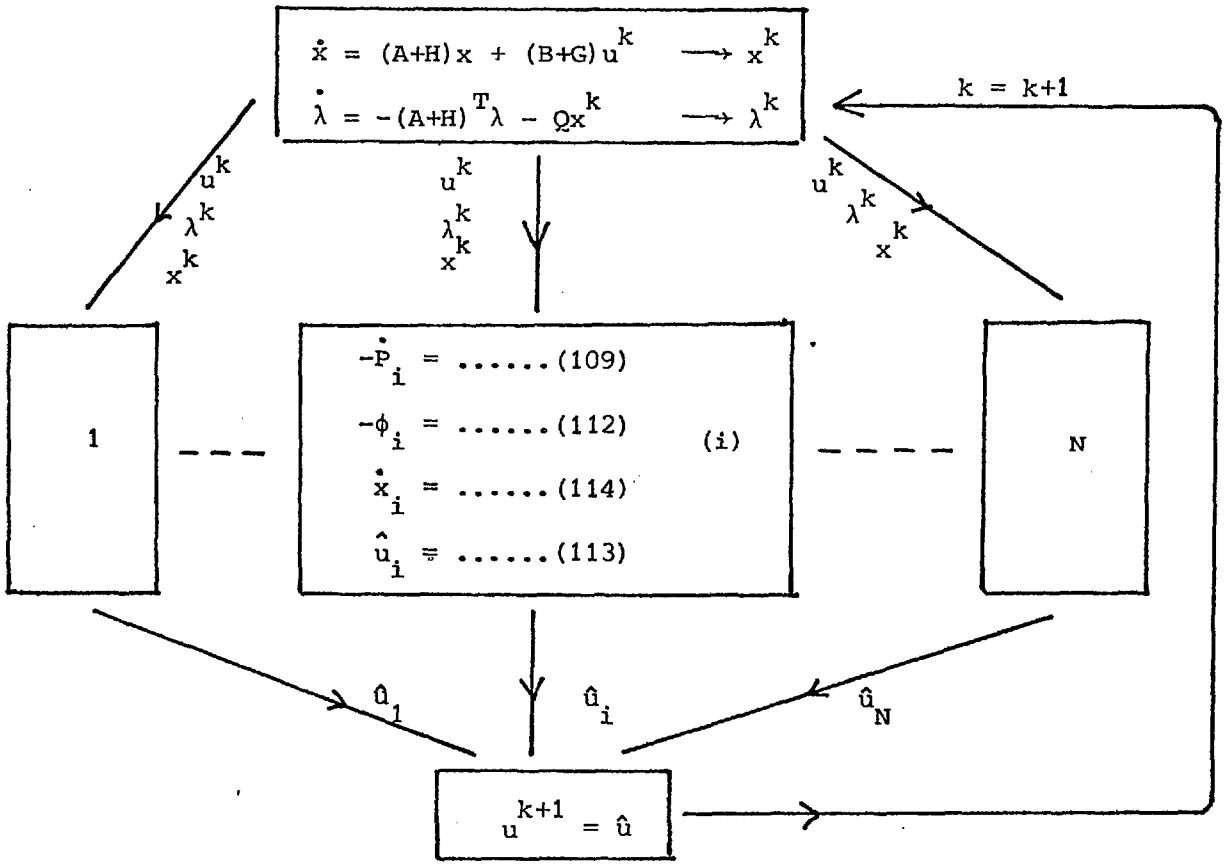
$$\dot{x}_i = (A_i - B_i R_i^{-1} B_i^T P_i)x_i - B_i R_i^{-1} B_i^T \phi_i - B_i R_i^{-1} \sum G_{ji}^T \lambda_j^k + \sum G_{ij} u_j \;;$$

$$+ \sum H_{ij} x_j^k \,, \qquad x_i(t_o) = x_{io} \;; \tag{114}$$

7) Calculate the local open loop controls by substituting the solution of (114) into (113).

8) Apply stopping condition ; if not met set k = k+1 and return to 3).

The flow of information in this procedure is as follows :



Identity between Interaction Prediction and Newton Raphson (K^{-1}) with $K = R+Q*QW$

We now show that this IP algorithm is identical to the NR(K^{-1}) algorithm discussed above. To do this, we look at the control sequences generated by the two algorithms.

We already have the IP control sequence, in (99) above . We now obtain the corresponding sequence for the NR(K^{-1}) algorithm. Recall the update formula

$$u^{k+1} = u^k - K^{-1}g(u^k) \tag{155}$$

where $$K = R + B^T\hat{\Phi}*Q\hat{\Phi}B \tag{156}$$

and $g(u^*) = [R+(B+G)^T\Phi^*Q\Phi(B+G)]u^k + (B+G)^T\Phi^*Qx_s$

or $g(u^*) = Lu^k + (B+G)^T\Phi^*Qx_s$.

Hence

$$u^{k+1} = (I - K^{-1}L)u^k - K^{-1}(B+G)^T\Phi^*Qx_s$$

$$= K^{-1}(K-L)u^k - K^{-1}(B+G)^T\Phi^*Qx_s$$

so that $u^{k+1} = (R+B^T\hat{\Phi}^*Q\hat{\Phi}B)^{-1}(B^T\hat{\Phi}^*Q\hat{\Phi}B - (B+G)^T\Phi^*Q\Phi(B+G))u^k$

$$- (R+B^T\hat{\Phi}^*Q\hat{\Phi}B)^{-1}(B+G)^T\hat{\Phi}Qx_s \cdot \qquad (158)$$

Again, as in the static case, we note that the inverted factor is the same in each update formula, (158) and (99), and it is a matter of lengthy and tedious algebra to show equivalence of the formulae.

Finally, we should like to relate the two proceduresgiven for *computing* the algorithm : the 'simple model' procedure used above, and the modification by interaction operator used here. In the present approach, the $NR(K^{-1})$ update

$$u^{k+1} = u^k - K^{-1}g(u^k) \qquad (159)$$

has been shown to be *equivalent* to the IP optimization problem :

$$\min_u \tfrac{1}{2}x^TQx + \tfrac{1}{2}u^TRu + \lambda^{kT}(Gu+Hx)$$

$$\text{s.t. } \dot{x} = Ax + Bu + z \qquad (160)$$

$$z = Gu^k + Hx^k$$

In the 'Simple model' approach, the form of the $NR(K^{-1})$ algorithm is respected, and the computation of $K^{-1}g$ *alone* was shown to be equivalent to the problem :

$$\min_h \tfrac{1}{2}x^TQx + \tfrac{1}{2}h^TRh - h^Tg$$

$$\text{s.t. } \dot{x} = Ax + Bh, \quad x(t_0)=0.$$

Hence in this new light the h^Tg term can be seen to be a *modification* of the subsystem criteria, to compensate for the loss of interaction on the system dynamics. However the two optimization problems are distinct.

It should be noted that the computational procedure (101)-(114) derived above is identical to that derived by Cohen, save for the difference that he allowed Q and R to be non-block diagonal. In our context this was a needless complication.

§5 Operator version of Takahara algorithm and multilevel schemes

We now derive operator analogues of the differential multilevel schemes, to see how they compare with the directly derived interaction prediction, and gradient schemes. Let us first establish the basic connections between the operator and differential description of the LQ regulator. In differential terms we have the following familiar scheme.

$$\min \tfrac{1}{2} \int_{t_o}^{t_f} (x^T Q x + u^T R u)\, dt \tag{121}$$

$$\text{s.t.} \quad \dot{x} = Ax + Bu\,, \quad x(t_o) = x_o\,. \tag{122}$$

The Hamiltonian is given by

$$H(x,u,\lambda,t) = \tfrac{1}{2}x^T Q x + \tfrac{1}{2}u^T R u + \lambda^T(Ax+Bu) \tag{123}$$

and the necessary conditions for optimality are

$$\begin{aligned} \frac{\partial H}{\partial \lambda} &= \dot{x} = Ax + Bu \\ \frac{\partial H}{\partial x} &= -\dot{\lambda} = Qx + A^T\lambda \\ \frac{\partial H}{\partial u} &= 0 = Ru + B^T\lambda \end{aligned} \tag{124}$$

$$\text{Hence } u^* = -R^{-1}B^T\lambda$$

$$\begin{aligned} \text{where } \dot{x} &= Ax - BR^{-1}B^T\lambda, \quad x(t_o) = x_o \\ \dot{\lambda} &= -Qx - A^T\lambda \quad , \quad \lambda(t_f) = 0\,. \end{aligned} \tag{125}$$

The corresponding operator scheme is

$$\min \tfrac{1}{2}\langle x, Qx\rangle + \tfrac{1}{2}\langle u, Ru\rangle \tag{126}$$

$$\text{s.t.} \quad x = \Phi Bu + x_s \tag{127}$$

$$\text{Hence } J(u) = G(u,P(u)) = \tfrac{1}{2}\langle x_s Q x_s\rangle + \tfrac{1}{2}\langle u, [R+B^T\Phi^T Q\Phi B]u\rangle + \langle u, B^T\Phi^T Q x_s\rangle \tag{128}$$

$$\therefore \quad \frac{dJ}{du}(u) = (Ru + B^T\Phi^T Q\Phi B)u + B^T\Phi^T Q x_s \tag{129}$$

$$= Ru + B^T\Phi^T Q x^u \tag{130}$$

$$\left.\begin{aligned} &\text{Hence } u^* = -R^{-1}B^T\lambda \\ &\text{where } \lambda = \Phi^T Q x^u \\ &\qquad x^u = x_s + \Phi Bu = x_s - \Phi BR^{-1}B^T\lambda \end{aligned}\right\} \tag{131}$$

In differential terms, (131) is (125).

The operator formulation allows more explicit manipulation (but no ready solution procedure via Riccati equations).

From (129),

$$u^* = -(R+B^T\Phi^T Q\Phi B)^{-1}B^T\Phi^T Q x_s \tag{132}$$

so

$$x^* = x_s + \Phi Bu^* \tag{133}$$

$$= x_s - \Phi B(R+B^T\Phi^T Q\Phi B)^{-1}B^T\Phi^T Q x_s \tag{134}$$

These are fundamental expressions showing open loop trajectories' dependence on initial state. We also have the 'feedback' forms :

$$u^* = -R^{-1}B^T\Phi^T Q(x_s + Bu^*)$$

$$\text{i.e.} \quad u^* = -(I + R^{-1}B^T\Phi^T Q\Phi B)^{-1}R^{-1}B^T\Phi^* Q x_s \tag{135}$$

We see readily that (135) is identical with (132), since

$$\begin{aligned} &(I + R^{-1}B^T\Phi^T Q\Phi B)^{-1}R^{-1} \\ = \;& (R^{-1}R + R^{-1}B^T\Phi^T Q\Phi B)^{-1}R^{-1} \\ = \;& (R + B^T\Phi^T Q\Phi B)^{-1}(R^{-1})^{-1}R^{-1} \\ = \;& (R + B^T\Phi^T Q\Phi B)^{-1}. \end{aligned}$$

We return now to the interconnected LQ regulator. In this case the 2PBVP to be solved is

$$\begin{aligned} \dot{x} &= (A+H)x - (B+G)R^{-1}(B+G)^T\lambda \quad , \quad x(t_o) = x_o \\ \dot{\lambda} &= -Qx - (A+H)^T\lambda \quad , \quad \lambda(t_f) = 0 \end{aligned} \tag{136}$$

The corresponding operator expression is

$$x = -\Phi(B+G)R^{-1}(B+G)^T\lambda + x_s$$

$$\lambda = \Phi^*Qx$$

so that

$$x = -\Phi(B+G)R^{-1}(B+G)^T\Phi^*Qx + x_s$$

or $$x = [I + \Phi(B+G)R^{-1}(B+G)^T\Phi^*Q]^{-1}x_s \tag{137}$$

which could also be found by substituting u* from (135) in (134). Similarly,

$$u^* = -R^{-1}(B+G)^T\Phi^*Q[I+\Phi(B+G)R^{-1}(B+G)^T\Phi^*Q]^{-1}x_s \tag{138}$$

which expresses dependence of optimal control on initial state.

Alternatively (136) may be considered in decoupled form :

$$\begin{aligned} \dot{x} &= Ax + Hx - (B+G)R^{-1}(B+G)^T\lambda \quad , \quad x(t_o) = x_o \\ \dot{\lambda} &= -Qx - A^T\lambda - H^T\lambda \quad , \quad \lambda(t_f) = 0 \end{aligned} \tag{139}$$

leading to the operator expression

$$\begin{aligned} x &= \hat{\Phi}Hx - \hat{\Phi}(B+G)R^{-1}(B+G)^T\lambda + \hat{x}_s \\ \lambda &= \hat{\Phi}^*Qx + \hat{\Phi}^*H^T\lambda \end{aligned} \tag{140}$$

or

$$\begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} \hat{\Phi}H & -\hat{\Phi}(B+G)R^{-1}(B+G)^T \\ \hat{\Phi}^*Q & \hat{\Phi}^*H^T \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} + \begin{bmatrix} \hat{x}_s \\ 0 \end{bmatrix} . \tag{141}$$

Expression (141) is nothing other than the operator analogue of the 2PBVP (136) which we solved, in Chapter 4, by a generalized successive approximations scheme. But the system (141) can also be solved by successive approximations : in this case it is a simple

linear system of the form

$$w = Aw + b$$

and so it is amenable, as was the static LQ regulator, to solution via the Jacobi iteration scheme. Setting

$$A = K + E,$$

we derive this scheme as

$$w^{k+1} = (I-K)^{-1}Ew^k + (I-K)^{-1}b.$$

Let us now choose K in such a way that we obtain an *operator analogue* of the well-known Takahara scheme.

Let us choose, for the system (141),

$$K = \begin{pmatrix} 0 & -\hat{\Phi}BR^{-1}B^T \\ \hat{\Phi}^*Q & 0 \end{pmatrix}, \quad E = \begin{pmatrix} \hat{\Phi}H & -\hat{\Phi}BR^{-1}G^T - \hat{\Phi}GR^{-1}(B+G)^T \\ 0 & \hat{\Phi}^*H^T \end{pmatrix}$$

.... (142)

We have distributed the system (141) as

$$\begin{pmatrix} x \\ \lambda \end{pmatrix}^{k+1} = \begin{pmatrix} 0 & -\hat{\Phi}BR^{-1}B^T \\ \hat{\Phi}^*Q & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix}^{k+1} + \begin{pmatrix} \hat{\Phi}H & -\hat{\Phi}BR^{-1}G^T - \hat{\Phi}GR^{-1}(B+G)^T \\ 0 & \hat{\Phi}^*H^T \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix}^k + \begin{pmatrix} \hat{x}_s \\ 0 \end{pmatrix}.$$

.... (143)

$$\text{i.e.} \quad x^{k+1} = -\hat{\Phi}BR^{-1}B^T\lambda^{k+1} + \hat{\Phi}Hx^k - [\hat{\Phi}BR^{-1}G^T + \hat{\Phi}GR^{-1}(B+G)^T]\lambda^k + \hat{x}_s$$

.... (144)

$$\lambda^{k+1} = \hat{\Phi}^*Qx^{k+1} + \hat{\Phi}^*H^T\lambda^k .$$

In differential terms (144) is equivalent to

$$\begin{aligned} \dot{x}^{k+1} &= Ax^{k+1} - BR^{-1}B^T\lambda^{k+1} + Hx^k - [BR^{-1}G^T + GR^{-1}(B+G)^T]\lambda^k, \quad x(t_o) = x_o \\ -\dot{\lambda}^{k+1} &= A^T\lambda^{k+1} + Qx^{k+1} + H^T\lambda^k, \quad \lambda(t_f) = 0 \end{aligned}$$

.... (145)

which is just the Takahara scheme as derived in Chapter 4, equation (28). Hence we have succeeded in giving an operator analogue of the Takahara scheme. The scheme (143) also gives us an alternative operator formulation of a convergence condition. This condition is

$$\| (I-K)^{-1}E \| < 1$$

i.e.
$$\left\| \begin{bmatrix} I & \hat{\Phi}BR^{-1}B^T \\ -\hat{\Phi}^*Q & I \end{bmatrix}^{-1} \begin{bmatrix} \hat{\Phi}H & -\hat{\Phi}BR^{-1}G^T - \hat{\Phi}GR^{-1}(B+G)^T \\ 0 & \hat{\Phi}^*H^T \end{bmatrix} \right\| < 1 \qquad (146)$$

Finally, as in the static case, let us discuss an optimization problem which is the equivalent of the Takahara scheme, to compare it with the Interaction Prediction approach.

Recall from Chapter 3 that we found that the equations of the Takahara scheme constituted the stationarity conditions of the scheme:

$$\min_u \int \tfrac{1}{2}x^TQx + \tfrac{1}{2}u^TRu + \lambda^{k^T}(Hx + Gu)dt \qquad (147)$$

$$\text{s.t.} \quad \dot{x} = Ax + Bu + z$$

$$z = Gu^k + Hx^k$$

Again it is striking how the *form* of the problem (147) is *identical* to the IP optimization problem : the difference once again is in the interpretation of u^k, x^k, λ^k. In the IP case, these are *system* variables, whereas in the Takahara scheme these are *local* variables computed at subsystem level. Hence once again we may interpret the Takahara scheme as a *local approximation* of the IP scheme.

Chapter 6

Concluding remarks

The contributions of the Thesis have been outlined in Chapter 1 § 3. The most important area for future work would seem to be the numerical application of the new algorithms to real-world interconnected systems.

- o -

Bibliography

[1] G. Kron, J. Appl Physics, 24, 995, 1953

[2] G. Dantzig & P. Wolfe, Op Res, 8, 101, 1960

[3] M.D. Mesarovic, D. Macko, Y. Takahara, Proc 4th IFAC World Conf., Poland, 1969

[4] L. Lasdon, Duality, and decomposition in mathematical programming, IEEE Trans Systems Sci. & Cyb., SSC-4,2, 86, 1968

[5] J. Pearson, Multilevel control systems, Proc 2nd IFAC Symp on Theory of Self-Adaptive Control Systems, 1965

[6] G. Cohen, On an algorithm of decentralized optimal control, J. Math Anal & Appl., 59, 2, 1977

[7] G. Cohen and P. Bernhard, Commande optimale, décentralization et jeux dynamiques, Dunod 1976

[8] M. Singh & A. Titli, Systems: decomposition, optimisation and control Pergamon Press, 1978

[9] L. Hakkala & J. Hirvonen, A hierarchical approach to state space constrained optimization problems, Int. J. Sys. Sci.

[10] M. Javdan & R. Richards, Decentralized control systems theory : a critical evaluation, Int. J. Control, 26, 1, 129, 1977

[11] G. White & M. Simmons, Analysis of complex systems, Phil. Trans. R. Soc. London. A, 287, 405, 1977

[12] W. Findeisen, Parametric optimization by primal method in multilevel systems, IEEE Trans. Systems Sci. Cyb., SSC-4, 2, 155, 1966

[13] M. Hestenes, JOTA, 4, 303, 1969

[15] W. Findeisen et al, On-line hierarchical control for steady-state systems, IEEE Trans Auto. Control, AC-23,2, 189, April 1978

[16] A.J. Fossard et al, Aperçus sur la commande hiérarchisée RAIRO, J-3, 3-40, August 1972

[17] R. Kulikowski et al, Optimization and control problems in large-scale systems, Proc 6th IFAC Symposium, vol III, 1976

[18] R. Kulikowski, Optimization of large-scale systems, Automatica 6, 315-331, 1970

[19] P. Tatjewski, A penalty function approach to coordination in multilevel optimization problems, RAIRO Automatique Systems Analysis, and Control, 12,3, 221-244, 1978

[20] M.D. Mesarovic, D. Macko & Y. Takahara, Theory of hierarchical multilevel systems, Academic Press, 1970

[21] A.M. Geoffrion, Elements of large-scale mathematical programming, Management Sci. Theory, 16, 652-691, 1970

[22] A.V. Fiacco & G.P. McCormick, Non-linear programming : sequential unconstrained minimization techniques, Wiley, New York, 1968

[23] D.P. Bertsekas, Multiplier methods : a survey, Automatica, 12, 133-145, 1976

[24] J.D. Pearson, Dynamic decomposition techniques, in D.A. Wismer (ed.), Optimization methods for large-scale systems with application, Academic Press, New York, 1971

[25] G. Cohen, Optimization by decomposition and coordination : a unified approach, IEEE Trans. Auto. Control, AC-23, 2, 222 April 1978

[26] M.S. Mahmoud, Multilevel systems control and applications : a survey, IEEE Trans Systems, Man, Cybernatics, SMC-7,3,125, March 1977

[27] M.D. Simmons & G.W.T. White, The decentralized profit maximization of interconnected production system, Part I,II, Int. J. Control, 25,5, 705, 1977

[28] V.I. Opoitsev, Dynamics of collective behaviour I homogenic systems, Automation and Remote Control, 4, p 154, April 1974; II Systems with bounded interelement interaction, ARC, 6, p 133, June 1974; III Heterogenic systems, ARC, 2, p 124, January 1975

[29] A.V. Malishevskii, Models of joint operation of many goal-oriented elements, I, Automation and Remote Control, 11, p 92, November 1972; II, ARC, 12, p 108, December 1972

[30] J.D. Pearson, Decomposition, coordination, and multilevel systems, IEEE Trans Systems Science, Cybernetics, SSC-2, 1, 36, August 1966

[31] Y. Takahara, Multilevel approach to dynamic optimization, Report SRC-59-A-64-21, Systems Research Center, Case Institute of Technology, Cleveland, Ohio, 1964

[32] M.S. Mahmord, Dynamic feedback methodology for interconnected control systems, Int. J. Control, 29, 5, 881, 1979

[33] M.G. Singh, M.F. Hassan & A. Titli, Multilevel feedback control for interconnected dynamical systems using the prediction principle, IEEE Trans Systems, Man, Cybernetics, SMC-6, 4, 233, April 1976

[34] L. Hakkala & J. Hirvonen, On coordination strategies for the interaction prediction principle using gradient and multiplier methods, Int. J. Systems Sci., 0.0.0. 1978

[35] J.D. Pearson, S. Reich, Decomposition of large optimal control problems, Proc IEE, 114, 6, 845, June 1967

[36] E.J. Bauman, C.T. Leondes, D.A. Wismer , Two-level optimization techniques for dynamic systems, Int. J. Control, 8,5, 473, 1968

[37] G. Grateloup & A. Titli, Two-level dynmaic optimization methods, JOTA, 15, 5, 533, 1975

[38] Y. Takahara & M.D. Mesarovic, Coordinability of dynamic systems, IEEE Trans.Auto.Systems, AC-14, 6, 688, December 1969

[39] M.D. Mesarovic, D. Macko & Y. Takahara, Two coordination principles and their application in large-scale systems control, Automatica, 6, 261, 1970

[40] M.G. Singh & M. Hassan, A comparison of two hierarchical optimization methods, Int J Systems Sci., 7,6, 603, 1976

[41] G. Cohen, A, Benveniste P. Bernhard, Commande hiérarchisée avec coordination en ligue d'un systéme dynamique, Revue RAIRO, J4,71, December 1972

[42] L.A. Hakkala, On infeasible gradient-type coordination algorithms for dynamical systems, Int J Systems Sci., 8, 5, 489, 1977

[43] G. Cohen, A min-max approach to closed-loop decentralized control, Ri erche di Automatica, 8, 87, April 1977

[44] Y. Nishikawa, T. Ojika, H. Shimazutsu & M. Okudaira, An interaction-coordination algorithm with modified performance index for non-linear optimal control problems, JOTA, 25, 2, 229, January 1978

[45] M.G. Singh, A new algorithm for the on-line multilevel control of large interconnected systems with fast dynamics, Int J Control, 21, 4, 587, 1975

[46] M.A. Hestenes, The conjugate gradient method for solving linear systems, in Proc Symp. on Applied Maths. vol VI, Num Anal. An Math Soc., p 83-105

[47] J.C. Allwright, Improving the conditioning of optimal control problems using simple models, 1972

[48] J.C. Allwright, Generalised conjugate-gradient algorithm for control optimization using models, Proc IEEE Conf. on Decision and Control, New Orleans, December 1972

[49] E.A. Freeman, On the optimization of linear, time invariant multivariant control systems using the contraction mapping principle, JOTA, 3(6), 416-443, 1969

[50] E.A. Freeman, Optimization of contraction-mapping algorithm for calculating optimal controls, Proc IEE, 119(9), 1365-1371, 1972

[51] J.C. Allwright, Contraction mapping algorithm with guaranteed convergence, Proc 5th IFAC World Congress, June 1972

[52] D.Q. Mayne, Decentralized control and large scale systems, Research Report, 75/25, Department of Computing and Control, Imperial College, 1975

[53] M.G. Singh, S.A.W. Drew & J.F. Coales, Comparisons of practical hierarchical control methods for interconnected dynamical systems, Automatica, 11, 331, 1975

[54] M.S. Mahmoud, A quantative comparison between two decentralized control approaches, Int J Control, 28, 2, 261, 1978

[55] F. Riesz & B. Sz-Nagy, Functional Analysis, Frederick Ungar New York, 1965

[56] M.S. Mahmoud, W.G. Vogt & M.H. Mickle, Multilevel control and optimization using generalized gradients technique, Int J. Control, 25, 4, 525, 1977

[57] J.D. Pearson, in D.A. Wismer (ed.), Optimization methods for large-scale systems with applications, Academic Press, New York, 1971

[58] D.M. Young, Iterative solution of large linear systems, Academic Press, 1971

[59] S.G. Mikhlin & K.L. Smolitskiy, Approximate methods for solution of differential and integral equations, American Elsevier, 1967

[60] M.A. Kranosel'skii et al , Approximate solution of operator equations, Wolters-Noordhoff, Groningen, 1972

[61] D.K. Faddeev & V.N. Faddeeva, Computational methods of linear algebra , W.H. Freeman, San Francisco, 1963

[62] F.R. Gantmacher , The theory of matrices, vol 1, Chelsea Publishing, New York, 1959

[63] J.H. Wilkinson, The algebraic eigenvalue problem, Clarendon Press, Oxford, 1965