A WEIGHTING FUNCTION APPROACH TO LINEAR CONTROL SYSTEMS DESIGN

by

DAVID HAROLD MEE

A thesis submitted for the degree of
Doctor of Philosophy

July 1969

Centre for Computing and Automation
Imperial College of Science and Technology
University of London.

ABSTRACT

This thesis adopts the approach that a linear system is specified in terms of its transfer operator or weighting function. Using this basic assumption, certain topics in the theory of linear control systems are investigated and some new results are presented.

An abstract point of view is followed wherever possible, to enable the tools of functional analysis to be applied. A brief review of linear systems theory is first given and certain concepts of control theory are shown to have abstract interpretations. The theory of optimality to a quadratic performance criterion is examined, and some new theoretical results and computational algorithms are given. The optimal control of stochastic processes is seen as a further example of the abstract theory. Filtering is shown to be a dual problem to control. Using the weighting function approach, estimation becomes a linear problem, and an algorithm is presented for solving this.

Modelling errors and their effect on control calculations is investigated. The insight that the weighting function method gives, enables results in sensitivity and cost bounds to be derived. A method of design which accounts for modelling errors 'a priori' is presented.

The final part of the thesis considers the problem of optimisation of the gains of a closed loop system when only incomplete state feedback is used. The insight provided by the state space approach is used to derive algorithms for the sub-optimal case, though these are formulated abstractly.

ACKNOWLEDGEMENTS

The financial support of the Commonwealth Scholarship Scheme has enabled me to live and study in England, and to produce this thesis. I would like to express my appreciation to the Commonwealth Scolarship Commission, their secretariat at the Association of Commonwealth Universities, and the British Council, for their kind and generous assistance.

I am also grateful for the help I have received from my supervisor Mr.G.F.Bryant, and for the many useful discussions I have had with members of the Control Group at Imperial College.

D.H.Mee

22 July 1969

CONTENTS

INTRODUCTION

During the past decade, the concept of "state" of a dynamical system has both dominated and directed control systems research. The state-space approach is especially useful for the study of optimal control problems, since the dynamic programming method of solution indicates that the optimal control is a (time-varying) function of the instantaneous state. In particular, the study of linear recursive models, with quadratic cost criteria leads to an optimal control problem whose solution is a linear function of state. The associated matrix Riccati equation and its solution are well known. The dual problem of the Kalman-Bucy filter has also been an important example of the power of the state-space approach.

Unfortunately, there are some fundamental practical drawbacks to the use of these methods in engineering design. The most usual complaint is that not all the states of a system are accessible to physical measurement, and hence unavailable for feedback control. This is not an insuperable difficulty, however, for the Kalman filter, or other kinds of observers, such as those proposed by Luenberger [L1], may be used to estimate the inaccessible states from the measured outputs, leading to compensator design.

A more serious difficulty is that control system designs based on optimality tend to become very complicated if the dimension of the state-space is large, and hence expensive to implement. This is decidedly non-optimal in terms of monetary value. The optimal design may depend critically on the structural details of the dynamical plant model. If the model is only an approximate representation of the true physical plant, it is then quite possible for the implemented control scheme to be sensitive to "minor" structural inaccuracies, and produce closed loop plant behaviour very different to that of the model.

All these disadvantages contribute to the fact that modern control theory has not had a great impact on industrial control design. However, classical linear control theory, based on transfer functions, Bode plots and Nyquist diagrams, still proves a very useful tool in control design, though the techniques become unwieldy for multivariable systems analysis. These methods show at a glance the effects due to the introduction of extra modelling dynamics, small time delays, and other non-dominant phenomena, which may produce large structural changes in a state-space model. This provides the motivation for our thesis, which follows an approach quite different from the standard state-space theory. In order to reduce all linear systems to a uniform description and a uniform level of difficulty, we have adopted the view that a linear system is represented by its impulse response, or weighting function. To be more general, we use the concept of a system operator which maps inputs into outputs. For time-invariant continuous systems, this reduces to a convolution integral with a pre-measured weighting function. Once the system operator is known, the transfer characteristics of the linear system are known, and in particular its behaviour in a feedback control scheme.

While this approach is a reversion to the original Wiener-Kolmogorov theory, many of the results and computational algorithms presented are original. In Chapter 1, the weighting function approach is abstracted to become a part of linear systems theory, and basic assumptions are re-appraised. The chapter sets out a summary of the theory of linear systems that we wish to use in the remainder of the thesis, and presents the notation and the abstract development that forms the framework for our theory. Apart from the method of approach, there is little original contribution in this first chapter, except for the section on stationarity and causality, which is formalised to fit in with the abstract development.

Chapter 2 develops the main body of results on optimality for a quadratic performance index. The author has made some theoretical and practical original contributions in this chapter. The proofs in Section 2.2 of the conditions of optimality are the author's own work. The discussion of the performance of gradient algorithms in relation to non-minimum phase is also new. Most of Sections 2.5, 2.6 and 2.7, which develop the theory of optimal control, is original, and new methods of spectral factorisation are given in Section 2.9. Compensator design is discussed in Section 2.9, and new design methods are presented, leading to a practical implementation of optimal control in a feedback structure. Properties of optimal systems are discussed in Section 2.11 and the author's generalisations of previous results are proved. The whole chapter tries to show what parts of optimality theory depend purely on the property of linearity, as distinct from a state-space representation.

The abstract theory can be used to derive results for stochastic processes, and its power is demonstrated in Chapter 3. Although most of the results are well known, the derivations, and approach, is original. However, Section 3.4, on identification and estimation, presents a new estimation result based on the maximum likelihood technique.

The remainder of the thesis is concerned with the design of sub-optimal control systems. Chapter 4 considers some practical problems associated with modelling, and optimal control theory as a design method for a class of systems. The power of the weighting function approach enables an original generalisation of results on system sensitivity to be derived. Other original results in this section include a stability bound, and the discussion of an "a priori" design method. Further research work is still required in these topics, which constitute a particularly difficult field.

Chapter 5 considers the problem of optimisation of a fixed feedback structure. New algorithms are developed for this problem, which are based on optimal control theory. Used in conjunction with some of the lower bound results of Chapter 4, computer-aided feedback control design techniques for multi-input/multi-output systems described in this chapter are considered to be of engineering utility. Virtually all of this chapter is original, and design examples are given.

The concluding sections of the thesis indicate further avenues of research, especially in the field of non-linear systems and their representation. While several new results have been demonstrated, many more new problems are raised, and these are summarised at the end of the thesis.

CONTRIBUTIONS OF THIS THESIS

The introduction has indicated the contents of this thesis, and in what respect it embodies the results of the author's own research and observation. These investigations appear to the author to advance the study of control systems in the following ways.

1. The abstract approach to linear systems, and its interpretations, lead to a greater understanding of the fundamental problems.

2. Convenient computational algorithms for optimal control and compensator design, when only empirical impulse responses are known, form a useful tool for control systems design and digital realisation.

3. Approximation and the related design methods of Chapter 4 give a guide to the limitations on control systems achievement, which are imposed by modelling. This limitation is not generally appreciated.

4. The fixed configuration design methods of the final chapter provide a new, and potentially useful, method of linear systems design. With further research, these methods may also find application in non-linear systems.

CHAPTER 1

SYSTEMS THEORY

1.1 Physical Systems and Mathematical Models

Time is a basic entity of the physical world. Measured quantities change their values as time advances and this evolution of physical systems is of interest to the control engineer, especially if the evolution can be influenced by external manipulations. We seek to characterise some physical processes by mathematical models, so as to be able to deduce their evolution, or to deduce the form of manipulations which will allow the system to perform in a desirable way.

When one starts with the premise that any mathematical model, no matter how complicated, can only be an approximate representation of a physical system, it is then easy to see that the one physical system may be modelled by many mathematical equations, with more or less accuracy. The model chosen should be complicated enough to portray the physical phenomena that is of interest, but no more so. Any further detail only confuses the issue, and makes computation difficult. However, it is essential that an estimate of the approximation errors be known. There is always the danger of using a simple model for a situation in which the simplifying assumptions are invalid, and this can very easily be done in control systems design, as will be pointed out in this thesis.

The simplest of models proposed for the representation of physical systems is the system of simultaneous algebraic equations. This model is not usually general enough for control system studies, which are concerned with dynamic systems, and transient phenomena. The complexity can be increased by introducing terms which depend on time rates of change of the variables, to give a set of differential equations, together with appropriate boundary conditions.

If these equations only involve a finite number of variables, then the model is represented by ordinary differential equations and we are dealing with lumped systems. On the other hand, the variables may be collections of functions of spatial coordinates, together with various spatial derivatives. We are then considering distributed systems and partial differential equations, and the boundary conditions may become quite complicated. The boundary conditions usually depend on the spatial geometry of the dynamic system. If the forcing or control signals appear in the dynamic equations, we have spatial control, but if they appear in the boundary conditions, then we have boundary control.

However, even partial differential equations do not exhaust the possible mathematical models. In particular, variables at one instant of time may depend on the values that they had at previous instants, leading to differential-difference equations.

Integro-differential equations, integral equations and other forms of functional equations are also proposed as system models, and many examples may be seen in the literature. [BC 1].

A further degree of complexity is added when we consider the variables in our equations to be random, with underlying probability distributions, and the physical systems generate stochastic processes. It is also possible to sample all these continuous time systems, both deterministic and stochastic, to obtain sequences of a discrete variable. These systems will be called discrete time systems, or sampled data systems.

The theory of functional analysis can be fruitfully applied to the design of engineering systems. Because of the abstract notation and concepts, many diverse engineering problems can be reduced to the same problem in functional analysis. In the rest of this chapter a brief outline of some of the important results of functional analysis is presented. Certain basic concepts of control theory are generalised and restated in the language of functional analysis to provide insight and a useful abstract framework.

1.2 Linear Algebra

Two important concepts in control engineering are "signal" and "system". In this section, we consider abstractions of these ideas, which have certain mathematical properties corresponding to the physical situation. Basic terminology and notation are introduced.

Consider first a set of underlying scalars which lie in a field $\mathcal{F}$ [H1; p.1], [BM 1; p.33]. In particular, the field is required to be ordered and complete, and will be taken to be the set of real numbers in this thesis. Signals are then considered to be elements of certain abstract linear vector spaces over $\mathcal{F}$ [H 1; p.3], [BM 1; p.52]. These spaces may be spanned by a set of linearly independent vectors which constitute a basis [H 1; p.7-13]. The cardinal number of this basis is the dimension of the space.

Systems will be defined as transformations from one space of signals to another (or possibly the same) space. If for all $u \in \mathcal{D} \subseteq \mathcal{U}$, $y \in \mathcal{R} \subseteq \mathcal{Y}$, where $\mathcal{U}, \mathcal{Y}$ are vector spaces,

$$y = T(u) \qquad 1.1.2$$

we call $\mathcal{D}$ the domain of T, and $\mathcal{R}$ its range, and write

$$T : \mathcal{D} \rightarrow \mathcal{R} \quad . \qquad 1.2.2$$

The system T is linear if its domain is a space $\mathcal{U}$, and if for all $u_1, u_2 \in \mathcal{U}$, and α_1, α_2 scalars,

$$T(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 T(u_1) + \alpha_2 T(u_2), \qquad 1.2.3$$

from which it is seen that the range of T, written $\mathcal{R}(T)$ is also a linear space. There may be a set of vectors $u \in \mathcal{U}$ for which $T(u) = 0$, and this set is denoted by $\mathcal{N}(T)$. This set is also a linear space and is called the null space of T. For linear transformations the brackets will be omitted, and we will write

$$y = T u.$$

Consider the class of all linear operators $\mathcal{T}_{uy}$, where if $T \in \mathcal{T}_{uy}$

$$T : \mathcal{U} \rightarrow \mathcal{Y}.$$

$\mathcal{T}_{uy}$ is then a vector space. For the space of linear operators $\mathcal{T}_{uu}$ a multiplication of operators can be defined. If $u \in \mathcal{U}$, and T_1, $T_2 \in \mathcal{T}_{uu}$, then

$$T_2(T_1 u) = (T_2 T_1)u = T_3 u ,$$

where T_3 is a new operator which maps $\mathcal{U}$ into itself, and we write

$$T_3 = T_2 T_1 .$$

This multiplication is associative, and distributive with respect to addition and so we have proved the following:

Theorem 1.2.1: The space of linear transformations from a vector space into itself under the operations of addition and multiplication as defined above form a linear algebra over the field $\mathcal{F}$. In general, neither commutativity of multiplication, nor existence of inverses are assumed.

A particular class of linear operators on a space $\mathcal{U}$ is the class of linear functionals, which map $\mathcal{U}$ into the scalar field $\mathcal{F}$. These functionals lie in a vector space over $\mathcal{F}$, called the algebraic conjugate or algebraic dual space $\mathcal{U}^f$ [T 1; p.34]. If $f \in \mathcal{U}^f$, and $u \in \mathcal{U}$ the special notation $<f, u>$ is used for the scalar function. We can also consider the space of functions $\mathcal{U}^{ff}$ on the space of functionals $\mathcal{U}^f$. If $\mathcal{U}^{ff} \equiv \mathcal{U}$, then $\mathcal{U}$ is called algebraically reflexive.

Theorem 1.2.2: [T 1; p.45] A space is algebraically reflexive if and only if it is a finite dimensional.

If $u \in \mathcal{U}$ and $y \in \mathcal{Y}$, and $\exists$ T, such that

$$y = T u ,$$

then T induces a linear transformation from $\mathcal{Y}^f$ into $\mathcal{U}^f$ denoted by T^*. To each $y^1 \in \mathcal{Y}^f$ there corresponds a $u^1 \in \mathcal{U}^f$, such that

$$<u^1, u> = <y^1, y> = <y^1, T u> \quad \text{for all } u \in \mathcal{U}$$

i.e.

$$u^1 = T^* y^1 \qquad 1.2.4$$

where T^* is the algebraic conjugate of T, and we have the identity

$$<y^1, T u> = <T^* y^1, u>. \qquad 1.2.5$$

The equation

$$y = T u$$

is also open to the following interpretation: since y, u, T are all members of vector spaces, we can consider the map

$$u : T \rightarrow y \, .$$

That is, u is considered to be a linear transformation on the space $\mathcal{T}$, mapping into $\mathcal{Y}$. Hence, if $\mathcal{T}^f$ is the dual space of $\mathcal{T}$, the transformation u induces another transformation u^* from $\mathcal{Y}^f$ to $\mathcal{T}^f$ such that

$$< y^1, T\, u > \; = \; < y^1 u^*, T > \, . \qquad 1.2.6$$

The chain of relationships for linear functionals become

$$< y^1, T\, u > \; = \; < T^* y^1, u > \; = \; < T^*, u\, y^{1*} > \; = \; < y^1 u^*, T > \qquad 1.2.7$$

This interpretation enables the control concept of correlation to be incorporated into the abstract framework. If there exists a one-to-one linear correspondence between the elements of two vector spaces, they are called isomorphic. [H 1; p.14].

It is often possible to consider composite vector spaces made up of simpler vector spaces. There are two main methods of combining spaces. The first is the direct sum of spaces (or sometimes called

Cartesian product). If $y_i \in \mathcal{Y}_i$ then

$$\mathcal{Y} = \mathcal{Y}_1 \oplus \mathcal{Y}_2 \oplus \cdots \oplus \mathcal{Y}_n$$

where $\mathcal{Y}$ is the space of ordered collections of the vectors $\{y_i\}$. [H 1; p.28]. A more complicated method of combination is the direct product, tensor product, Kronecker product, or outer product, denoted by

$$\mathcal{U} = \mathcal{U}_1 \otimes \mathcal{U}_2 \otimes \cdots \otimes \mathcal{U}_n .$$

The tensor product space $\mathcal{U}_1 \otimes \mathcal{U}_2$ is defined to be the set of finite formal sums

$$\sum_i a_i \otimes b_i$$

where $a_i \in \mathcal{U}_1$, and $b_i \in \mathcal{U}_2$, and the operation $a_i \otimes b_i$ is bilinear. [H 1; p.40], [BM 1; p.187], [DS 1; p.90].

If the set $\{u_i^1\}$ is a basis of $\mathcal{U}_1$ and $\{u_j^2\}$ is a basis of $\mathcal{U}_2$, then $\{u_i^1 \times u_j^2\}$ forms a basis of $\mathcal{U}_1 \otimes \mathcal{U}_2$. [H 1; p.40]. In the final chapter of this thesis, we shall be particularly interested in transformations from product spaces to the elemental spaces which are called contractions.

<u>Example 1.2.1</u>: Consider $\mathcal{U}_1$ = space of 3-dim coordinate vectors, and $\mathcal{U}_2$ = space of 2-dim coordinate vectors. Then $\mathcal{U}_3 = \mathcal{U}_1 \oplus \mathcal{U}_2$ in the 5-dim space of coordinate vectors, whereas $\mathcal{U}_4 = \mathcal{U}_1 \otimes \mathcal{U}_2$ is the 6-dim space of 2 x 3 matrices, corresponding to the space

of sums of 2×3 dyads. The product $a \otimes b$ corresponds to $a\,b^T$.

For finite dimensional spaces, we have the following relations [H 1; p.30, p.37].

Theorem 1.2.3: If $\mathcal{U} = \mathcal{U}_1 \oplus \mathcal{U}_2$, $\dim \mathcal{U} = \dim \mathcal{U}_1 + \dim \mathcal{U}_2$.

Theorem 1.2.4: If $\mathcal{U} = \mathcal{U}_1 \otimes \mathcal{U}_2$, $\dim \mathcal{U} = \dim \mathcal{U}_1 \times \dim \mathcal{U}_2$.

1.3 Linear Control Systems and Linear Spaces

Using control systems terminology, we shall consider the dependent vectors or signals as system outputs, and the independent variables as system inputs. Inputs may be of two kinds; those which can be directly manipulated, and those which cannot. The former are called controls and the latter disturbances. If u represents controls, and v disturbances, then the output y is given by

$$y = T(u, v) \qquad 1.3.1$$

and for non-linear systems

$$y \neq T(u, o) + T(o, v)$$

in general. However, for linear systems the law of superposition holds.

i.e. $$y = y_1 + y_o \qquad 1.3.2$$

where y_o represents the output of the system due to disturbances alone, while y_1 represents the output due to controls. In fact

$$y_1 = W u$$

where W is a linear transformation on the space of controls and so the general explicit linear system is given by

$$y = y_o + W u \; . \qquad 1.3.3$$

These representations are explicit, whereas the general physical relationships usually derived are of the implicit type;

$$\left.\begin{array}{c} L y = B u \\ + \text{ boundary conditions} \end{array}\right\} \qquad 1.3.4$$

where L and B are linear operators. We assume that the form 1.3.3 is derivable from 1.3.4 and our results will usually be stated in terms of the form 1.3.3. In fact, the title of the thesis stems from representation 1.3.3, since in many cases W is representable by an integral kernel or weighting function.

Because of its abstract nature, the representation 1.3.3 can cover many situations. In the rest of this section, some examples of linear spaces of signals are considered.

For single input-single output systems, input and output can be deterministic functions of time. There are many kinds of functions of a single variable, and these may be defined over different intervals. Common spaces are;

(a) The spaces $\mathcal{C}^n(T)$, consisting of the set of functions of a single variable $t \in \{T\}$ where T is a point set, and all of whose derivatives up to n:th order exist and are continuous. Addition and scalar multiplication are defined in the usual way.

(b) The spaces $L_p (T)$, consisting of the equivalence classes of functions which are identical almost everywhere and which are absolutely p:th power Lebesque integrable over the point set $\{T\}$ with respect to interval measure. $(p \geqq 1)$.

(c) The sequence l_p $(p \geqq 1)$, consisting of sequences of points (finite or infinite) which are absolutely p:th-power summable. These spaces are very useful for sampled-data system studies.

(d) The spaces of n:th order distributions whose elements can be considered to be n:th-order δ-functions and all smoother functions. [K 1; p.265]. These have a special duality relationship with the spaces (a).

The above examples comprise the spaces of functions of one variable that will be considered in this thesis. Dunford and Schwartz [DS 1] present a comprehensive list of function spaces. For a fixed value of the time variable, the signal may not be a scalar, as in single input - single output deterministic systems, but an element of another vector space. In particular for multivariable systems this space may be (e) Euclidean n-dimensional vector space, over the real field denoted by R^n.

Using the direct product notation, these multivariable signals are considered vectors in the space $R^n \otimes F$, where F is a space of time functions.

Signals can be functions of two or more variables, and these occur in distributed systems, when functions are defined over spatial domains. Suitable generalisations of the spaces (a)-(d) are then used. We would like to be able to use the algebraic notions of direct sum and product to form these more complicated spaces. However, the direct product of two infinite dimensional spaces is difficult to define usefully. [DS 1; p.90]. We shall consider this question further in the next section.

1.4 Topologies, Metrics and Norms

To proceed with problems of interest, it is necessary to introduce analytical or geometrical considerations into the discussion so that concepts of distance, convergence, continuity, etc., can be used. In this section, a brief summary of the basic logical development of functional analysis is presented, to lead to the most important useful spaces, the Banach and Hilbert spaces.

In general, the notation and terminology of Taylor[T1]chapters 2 and 3, are followed. One begins by defining a topological space as consisting of certain sets called open sets, and their complements, the closed sets. [T 1; p.57]. From these notions, it is possible to define continuity of functions[T 1; p.61], and compactness of sets [T1; p.62]via the Heine-Borel condition of finite open coverings. An interesting property is that continuous functions map compact sets into compact sets. A set is called relatively compact if its closure is compact. Separability and separability axioms enable a classification of topological spaces.

A specialisation of these spaces is the metric space, on which a distance function is defined[T 1; p.68]. It is in fact possible to start with the distance function and from this define open sets. This method sets up a topology, called the metric topology, which enables the notions of convergence, continuity and compactness to be specialised to metric spaces. A set of spheres [T 1; p.68] can be constructed around any point, and a set is bounded if it is contained in some sphere. Metric spaces have some important properties [T 1; p.70, 71] among which are

Theorem 1.4.1: A compact set in a metric space is closed and bounded.

Theorem 1.4.2: In a metric space a set S is compact if every sequence in S contains a convergent subsequence with limit in S.

Theorem 1.4.3: A compact metric space is separable.

If there exists an invertable mapping f from one metric space X to another metric space Y, such that the distance between any two points in X and the distance between the image points under f in Y are identical, then X and Y are isometric.

The property of completeness is of importance[T 1; p.74] . This is the property possessed by a metric space X when every Cauchy sequence in X possess a limit in X. Incomplete metric spaces can be regarded as dense subsets in a complete metric space [T1; p.74].

When the topological and linear algebraic structures are combined we have a very rich field in which to investigate linear systems theory. Chapter 3 of Taylor [T 1] provides an introduction to the linear theory, and begins by defining a topological linear space as the combination of a linear vector space and a topolcgical space, where addition and scalar multiplication are continuous operations.

If linear properties are imposed on a metric, the concept of a norm can be defined. [T1; p.83]. If $x_1, x_2 \in X$ a normed linear space, then the norm of x_1 is denoted by $\| x_1 \|$, and the distance between x_1 and x_2 is $\| x_1 - x_2 \|$.

A normal linear space is called a Banach space if it is complete [T 1; p.98]. Banach spaces are the most usual kind of space of signals met with in control theory.

On some spaces, it may be possible to define a norm by means of an inner product, or a bilinear map of two vectors into the scalars. [T 1; p.106]. If X is an inner product space, and $x_1, x_2 \in X$, then their inner product is denoted by $< x_1, x_2 >$. The Schwartz and Bessel inequalities hold for inner product spaces, and the concept of orthogonality can be introduced. If the inner product space is complete it is called a Hilbert space, and these include the finite dimensional Euclidean spaces R^n, with the normal scalar product. With an appropriate norm, all the spaces mentioned in Section 1.3 are Banach spaces (e.g. L_p spaces are complete by the Riesz-Fisher theorem). However, the L_2 and l_2 spaces are the only simple infinite dimensional Hilbert spaces in our examples.

The concept of direct sum of Hilbert spaces causes no difficulty, and the sum space becomes a Hilbert space with an appropriate inner product. [DS 1; p.255-7]. We wish to define the direct product of two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, such that

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \qquad 1.4.1$$

is a Hilbert space. This can be done as follows. Define the space PH to be the space of finite sums of formal products $a \otimes b$.

i.e. $$x_1 = \sum_{i=1}^{N_1} a_i \otimes b_i \qquad 1.4.2$$

where $x_1 \in PH$, $a_i \in \mathcal{H}_1$ and $b_i \in \mathcal{H}_2$,

Also let

$$x_2 = \sum_{j=1}^{N_2} c_j \otimes d_j \, . \qquad 1.4.3$$

The $\otimes$ symbol implies a bilinear operation. A simple product $a \otimes b$ is called a dyad. In general, the expansions 1.4.2-3 are non unique. An inner product is defined on PH as

$$<x_1, x_2> = < \sum_{i=1}^{N_1} a_i \otimes b_i \, , \sum_{j=1}^{N_2} c_j \otimes d_j >$$

$$= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} < a_i \otimes b_i \, , c_j \otimes d_j >$$

$$= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} < a_i, c_j > \cdot < b_i, d_j > \, .$$

That this is a valid inner product is easily verified. We show that

$< x_1, x_1 > \geqq 0$ and

$< x_1, x_1 > \neq 0$ if $x \neq 0$.

Now $<x_1, x_1> = \sum_{i=1}^{N_1} \sum_{j=1}^{N_1} < a_i, a_j > < b_i, b_j > \cdot$

If $<a_i, a_j>$ is the i,j th element of the matrix A, and $<b_i, b_j>$ is the i,j th element of the matrix B, then

$$<x_1, x_1> \;=\; \text{tr } AB.$$

But A and B are positive semi-definite, being matrices, and so by the lemma 5.4.1 (to be proved in Chapter 5)

$$\begin{aligned} <x_1, x_1> &= \text{tr } AB \\ &\geqq 0 . \end{aligned}$$

We assume that it is possible to represent all vectors by the sum of orthogonal dyads.

Now if $\quad x_1 = a \otimes b$

then $\quad \begin{aligned} <x_1, x_1> &= <a \otimes b, a \otimes b> \\ &= <a , a> \; <b , b> , \\ &\neq 0 \end{aligned}$

unless either $a = 0$ or $b = 0$, in which case, since $a \otimes b$ is bilinear, $x_1 = 0$.

If $\quad x_1 = \sum_i d_i$

where $\quad < d_i, d_j > = 0 \quad$ for $i \neq j$

then $\quad < x_1, x_1 > = \sum_i <d_i, d_i>$

$$\neq 0$$

unless all $d_i = 0$.

Thus PH is a pre-Hilbert space, but it is not, in general, complete. We define $\mathcal{H}$ to be the completion [T 1; p.74] of PH so that $\mathcal{H}$ becomes a Hilbert space.

1.5 Linear Operators I

The effect of topologies on linear operators can now be discussed. One of the important topological notions is continuity. The continuous linear operators from a topological linear space X into a space Y form a subspace of all linear operators from X to Y. If X = Y, then they form a sub-algebra of the algebra of all linear operators from X into itself.

If X, Y are normed linear spaces then a linear operator A is continuous if, and only if, it is bounded; [T 1; p.162] i.e.

$$\| y \| = \| A x \| < \infty \qquad \forall \ \| x \| \leq 1 .$$

For such an A, we define

$$\| A \| = \sup_{\| x \| = 1} \| A x \| . \qquad 1.5.1$$

The space of continuous linear operators with this norm is a normed linear space, $\mathfrak{J}$. If Y is complete, then so is $\mathfrak{J}$. If X is a Banach space, then the set of continuous linear operators from X into itself is a Banach algebra.

As in the algebraic case, the special linear operators which map out a space onto the scalars are considered. However, in this case, the set of continuous linear functionals form a topological space X^* which is called the topological conjugate, or topological dual space of X, and is a subspace of X^f. If X is normed, so is X^*.

We may also consider the normed conjugate space X^{**} of X^*, and in some cases X^{**} is isomorphic and isometric (congruent) to X. In this case, X is norm-reflexive, and, in contrast to the purely algebraic case, infinite dimensional spaces can be norm-reflexive.

Also following the algebraic development, the continuous conjugate of a continuous linear operator [T1; p.213] is defined, and will be used extensively in the development. For the important case of Hilbert spaces X, Y, the dual space X^* is congruent to X, and linear functionals are identified as inner products, for which we use the same notation. Similarly, if A maps X into Y, then the adjoint operator A^* maps Y into X, and is defined by

$$< A^* y, x > = < y, A x >$$

In this way, the adjoint is identified with the continuous conjugate.

One of the most important properties of a Hilbert space is the orthogonal decomposition property. If M is a subspace of X, a Hilbert space, and $M^{\perp}$ is the space of all vectors y such that $\langle y, x \rangle = 0 \ \forall \ x \in M$, then $X = M \oplus M^{\perp}$.

The following is a summary of the properties of linear operators and their adjoints on Hilbert space [T 1; p.250-1][D S 1; p.480].

<u>Theorem 1.5.1</u> $\| A^* \| = \| A \|$ 1.5.2

$$(AB)^* = B^* A^* \qquad 1.5.3$$

$$A^{**} = A \qquad 1.5.4$$

$$\left\{ \overline{R(A)} \right\}^{\perp} = \mathcal{N}(A^*)$$

$$\overline{R(A)} = \left\{ \mathcal{N}(A^*) \right\}^{\perp}$$

$$\left\{ \overline{R(A^*)} \right\} = \mathcal{N}(A)$$

$$\overline{R(A^*)} = \left\{ \mathcal{N}(A) \right\}^{\perp}$$

If $A^* = A$, A is called self-adjoint, and this class of operators is important in control theory.

1.6 Linear Operators II

In this continuation of the summary of linear operator theory, we will assume that our basic spaces X, Y are normed. If $T: X \to Y$ then $T \in \mathcal{J}_{XY}$. Then we can state

Theorem 1.6.1: T^{-1} exists and is continuous if and only if there exists a constant $m > 0$, such that

$$m \| x \| \leqq \| T x \| \quad \forall \; x \in X .$$

Certainly a necessary condition for the existence of a bounded inverse is that $T x = 0 \implies x = 0$. However, this is not sufficient. For from Titchmarsh's theorem [M 1], convolution operations have this property, and yet may not have bounded inverses.

Example 1.6.1: Convolution of a function with the weighting function e^{-t} is a bounded operation in any L_p space. However, the inverse operation, which consists of adding the function to its derivative, is unbounded.

If T^{-1} does exist, then [T 1; p.250],

$$(T^{-1})^* = (T^*)^{-1} . \qquad 1.6.1$$

A common type of operator arising in control systems studies is the compact operator [T 1; p.274] or completely continuous operator. An operator is compact if it maps all closed bounded sets in its domain into compact sets in its range. The following general theorems are important:

Theorem 1.6.2: Schauder;[D&S1; p.485] A linear operator from one Banach space to another is compact if its adjoint is compact.

Theorem 1.6.3: A compact operator is continuous.

Theorem 1.6.4: If A is compact, and B is bounded, where $A, B \in \mathcal{J}_{XX}$, then $A B$, $B A$ are compact.

Theorem 1.6.5: A compact linear operator $T \in \mathcal{J}_{XX}$, does not have a bounded inverse if there exist non-compact closed and bounded sets in X.

Proof 1.6.5: (Contradiction)

Assume T^{-1} exists, bounded.

Then, if V is a bounded, closed, but non-compact set in space X, $T^{-1} V$ is a closed bounded set, since T^{-1}, being bounded, maps bounded sets into bounded sets. Moreover, since T is continuous, T^{-1} maps closed sets into closed sets. However, $T(T^{-1} V)$ is a compact set by assumption. But $T T^{-1} = I$.

Therefore V is a compact set, which is a contradiction. //

For T a compact operator on these spaces,

$$\|T x\| \not\geq m\|x\| \qquad \forall x \qquad \text{for any } m > 0.$$

An important class of compact operators are those corresponding to "smoothing" filters. For example, the operator given by

$$y(t) = T u(t) = \int_0^t W(t-\tau)\, u(\tau)\, d\tau \qquad t \in [o,T]$$

where $W(t) \in C[o, T]$, and $u(\cdot) \in L_2[o, T]$ is compact. This can be proved using Arzela's theorem on compact sets in the space of continuous functions. [V 1; p.314] . The function $y(t)$ is "smoother" than $u(t)$ in the sense that $y(t)$ is absolutely continuous.

Operators may be unbounded, but may have closed range [T 1; p.175]. These comprise operators such as differentiation on L_2, etc., which may have compact inverses.

A class of operators on Hilbert space into itself, that finds some application is the set of isometric transformations. An operator U is isometric if

$$< y, x > = < Uy, Ux > \qquad \forall\ x, y \in X ,$$

or equivalently

$$U^* U = I .$$

However, U^{-1} may not exist, in which case U^* is only a left inverse of U. If U^{-1} does exist, then

$$U^{-1} = U^*$$

and U is called a unitary transformation.

Theorem 1.6.6: In finite dimensional space, every isometric transformation is unitary. [RN 1; p.260].

Example 1.6.2: An isometric transformation which is not unitary is given by the shift operator T_1 on $l_2(o, \infty)$, defined by

$$T_1 : [u_o, u_1, u_2, \ldots] \longrightarrow [0, u_o, u_1, u_2, \ldots]$$

[RN 1; p.281]. T_1^* is defined $\forall\ x \in X$, (but T^{-1} is not).

Theorem 1.6.7: If there exists a W, such that $W^* W = P$, a self adjoint operator on Hilbert space, then the set of all such W differ from each other by an isometric transformation.

Proof: (a) Sufficiency: Let $W_1 = U W$.

If U is isometric, then

$$W_1^* W_1 = W^* U^* U W = W^* W.$$

(b) Necessity: Assume $W_2^* W_2 = P$.

Then $<x, Px> = <x, W_2^* W_2 x> = <W_2 x, W_2 x>$

$$= <y_2, y_2> .$$

Consider the transformation

$$y_2 = W_2 x .$$

We also have that

$$y = W x .$$

So this transformation defines another transformation U, such that

$$y_2 = U y = U W x .$$

Then $<y_2, y_2> = <U W x, U W x> = <U y, U y>$.

But $< x, P\,x > \;=\; < W\,x, W\,x > \;=\; < y, y >$.

Hence U is isometric.

Example 1.6.3: An example of an isometric transformation is the all-pass network of classical linear control theory. E.g.

$$U(s) = \frac{a - s}{a + s} \quad .$$

Then $$< U\,x, U\,y > = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{a + s}{a - s} \,.\, x(-s) \,.\, \frac{a - s}{a + s}\, y(s)\, ds$$

$$= \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} x(-s)\, y(s)\, ds$$

$$= < x, y > .$$

1.7 Representations of Linear Operators and Functionals

In this section, we investigate some particular spaces and examples of linear functionals and operators, with particular reference to linear control systems. The spaces L_p and l_p are of particular interest, and it is known that the normed conjugate of L_p (l_p) is L_q (l_q) for $p, q \geq 1$, and $\frac{1}{p} + \frac{1}{q} = 1$ [T 1; p.193, p.380]. Also if $1 < p < \infty$, then L_p (l_p) is norm reflexive. However, this is not so for $p = 1$ or ∞ . Also it is known that for the spaces $C^n(T)$

where T is compact, the normed conjugates are the spaces of n:th order distributions. [K 1; p.293] [H 2; Ch.4].

Linear operators can be often represented in much the same way as functionals. It is known that the most general operator from $L_p[0,1]$ to L_q, p, q >1, has the form

$$y(t) = T\,u(t) = \frac{d}{ds}\int_0^{+1} K(t, s)\, u(t)\, dt \qquad 1.7.1$$

[DS 1; p.490]. However, no satisfactory expression for the norm of T is known, and no conditions on K are known which are equivalent to the compactness of T. However, sufficient conditions for compactness of some operators are known. [V 1; p.313]. In control systems design, we are interested in operators of the form

$$T\,u = \int_o^t W(t, \tau)\, u(\tau)\, d\tau \qquad 1.7.2$$

or

$$T\,u = \int_{-\infty}^{\infty} W(t,\tau)\, u(\tau)\, d\tau \; . \qquad 1.7.3.$$

If $W(t, \tau)$ is just a function of $t - \tau$, then T is a convolution operator. If $W(t) \in L_1[o, \infty)$, and $u(t) \in L_p[o, \infty)$, then

$$y(t) = T\,u(t) \in L_p[o, \infty)$$

and

$$\|y\|_p \leqq \|W\|_1 \; \|u\|_p \qquad 1.7.4$$

[DS 1; p.528].

Adjoints are found from the formulae for linear operators, and linear functionals. For a map from $R^m \otimes L_2 [0, T]$ into $R^n \otimes L_2 [0, T]$ with the operator of 1.7.2, we have

$$T^* y(t) = \int_t^T W^T (\tau, t) y(\tau) d\tau \qquad 1.7.5$$

where W is a matrix valued function. This can be shown as follows:

$$\int_0^T y^T(t) (\int_0^t W(t, \tau) u(\tau) d\tau) dt$$

$$= \int_0^T \int_0^t y^T(t) W(t, \tau) u(\tau) d\tau dt$$

$$= \int_0^T d\tau \int_\tau^T y^T(t) W(t, \tau) u(\tau) dt$$

(where change of order of integration is permissible, which is certainly true if the integrand is continuous in the two dimensional region)

$$= \int_0^T dt \int_t^T y^T(\tau) W(\tau, t) u(t) d\tau$$

(changing roles of dummy variables)

$$= \int_0^T [\int_t^T W^T(\tau, t) y(\tau) d\tau]^T u(t) dt .$$

Another common operator arising in control theory is the discrete time weighting sequence. Instead of the above development, we replace integral signs by summations, and continuous variables by discrete, to obtain

$$(T^{*} y)_k = \sum_{i=k}^{N} W^{T}{}_{i,k} \, y_i \, . \qquad 1.7.6$$

Transforms from a time domain into a complex variable domain constitute an important type of linear operator. In general, these are invertable operations and hence set up isomorphisms between spaces. In much of our work, it matters little whether the formulae are interpreted in the time domain or the frequency (complex variable) domain. We shall freely use transform theory, when it simplifies the development.

1.8 Stationarity and Causality

The terms stationarity and causality are important physical concepts. However, in keeping with our preference for abstractions, we shall give meaning to these terms in a way which does not depend on coordinates.

To define stationarity or time-invariance of operators, we first consider the semi-group of operators called shift operators. These are a one-parameter family, and obey the relations

$$T(t_1 + t_2) = T(t_1)\, T(t_2) = T(t_2)\, T(t_1) \qquad 1.8.1$$

$$T(0) = I \qquad 1.8.2$$

$$\| T(t) \| = 1 \quad . \qquad 1.8.3$$

The parameter $t \in \{S\}$, and may be continuous or discrete. If $t \in (-\infty, \infty)$, then this set constitutes a full abelian group. However, it is only a semi-group on the half-line $[0, \infty)$, since then $T(t)$ has no inverse. An operator W is stationary, or time invariant if and only if it commutes with $T(t)$ for all t in its domain. i.e.

$$T_t W u = W T_t u \quad . \qquad 1.8.4$$

Theorem 1.8.1: The set of time-invariant operators form a sub-algebra of the algebra of operators $\mathcal{J}$ from a vector space into itself.

Proof: Time-invariant operators obviously form a vector space. Also, if V, U are time invariant, then

$$T_t V U u = V T_t (U v)$$

$$= V U T_t u \ .$$

i.e. $V U$ is also time-invariant.

Examples: The space of signals is the Hilbert space $L_2(-\infty, \infty)$. $T_t u$ is defined by

$$T_t : u(s) \longrightarrow u(s + t) \ .$$

Then the operator W defined by

$$y(t) = \int_{-\infty}^{\infty} W(t - \tau)\, u(\tau)\, d\tau$$

is time-invariant. For we have that

$$y(t+s) = \int_{-\infty}^{\infty} W(t+s-\tau)\, u(\tau)\, d\tau$$

Put $\xi = \tau - s$, giving

$$y(t+s) = \int_{-\infty}^{\infty} W(t-\xi)\, u(\xi+s)\, d\xi$$

i.e. $\quad WTu = TWu$.

The operator R defined by

$$y(t) = R(t)\, u(t)$$

is not time-invariant, unless $R(t)$ is a constant.

Causality is even more basic than stationarity in the modelling of physical systems. To define causality, we assume that it is possible to decompose the space of signals X into two fundamental subspaces M_+ and M_-, such that

$$X = M_+ \oplus M_- \qquad 1.8.6$$

and $$M_+ \cap M_- = 0 . \qquad 1.8.7$$

This decomposition defines two projections π_+ and π_- [T 1; p.240-42]

where $$\pi_- = I - \pi_+ \qquad 1.8.8$$

and $$\left.\begin{aligned} \pi_+^{\,2} &= \pi_+ \\ \pi_-^{\,2} &= I - 2\pi_+ + \pi_+^{\,2} = I - \pi_+ = \pi_- \end{aligned}\right\} 1.8.9$$

$$\pi_+ \pi_- = \pi_- \pi_+ = 0 . \qquad 1.8.10$$

$$\left.\begin{aligned} \mathcal{R}(\pi_+) &= \mathcal{N}(\pi_-) = M_+ . \\ \mathcal{R}(\pi_-) &= \mathcal{N}(\pi_+) = M_- . \end{aligned}\right. \qquad 1.8.11$$

Now since π_+, π_- are linear operators mapping X into itself, they belong to the linear algebra of operators from X into itself. Considering π_+, π_- in this latter interpretation, we define W to be causal, or non-anticipatory, if for all $u \in M_+$, $Wu \in M_+$. The set of causal operators is denoted by $\mathcal{J}_+$. Similarly $W \in \mathcal{J}_-$ if, $\forall\, u \in M_-$, $Wu \in M_-$, and W is then completely anticipatory.

<u>Theorem 1.8.2</u>: The set of causal operators form a sub-algebra $\mathcal{J}_+$ of the algebra $\mathcal{J}_X$ of linear operators from X into itself.

<u>Proof</u>: It is obvious that the causal operators form a vector space, since, if $T, S \in \mathcal{J}_+$, then so does

$$\alpha_1 T + \alpha_2 S \quad ; \quad \alpha_1, \ \alpha_2 \ \text{scalars.}$$

Also $T S \in \mathcal{J}_+$, since for $u \in M_+$

$$S u = y \in M_+ \quad .$$

But $\quad T y \in M_+$

i.e. $\quad T S u \in M_+ \quad \forall u \in M_+ \quad .$

So $\quad T S \in \mathcal{J}_+ .$

Examples: Once again, we choose the basic space as $L_2(-\infty , \infty)$.

We let $\quad M_+ = L_2 [o, \infty)$

and $\quad M_- = L_2 (-\infty ; o] \quad .$

Then we have $\quad X = M_+ \oplus M_-$

and $\quad M_+ \cap M_- = 0 \quad .$

If $u \in L_2[o, \infty)$, then the operator defined by

$$y(t) = \int_o^t W(t, \tau) u(\tau) d\tau$$

(where $W(t, \tau)$ is such that $y \in L_2[o, \infty))$, is causal. On the other hand

$$y(t) = \int_t^\infty W(\tau, t) u(\tau) d \tau$$

defines a completely anticipatory operator.

While $M_+ \cap M_-$ is void, $\mathcal{J}_+ \cap \mathcal{J}_-$ is not necessarily so. Consider the operator R defined by

$$y(t) = R(t)\,u(t) \qquad L_2(-\infty, \infty)$$

Then $R \in \mathcal{J}_+ \cap \mathcal{J}_-$.

In general, the subspaces M_+ and M_- are chosen such that the shift operator $T(t)$ is causal for $t > 0$, and non-causal for $t < 0$. Both causality and stationarity have been defined on mappings from a space into itself. These definitions can be easily generalised to mappings from one space to another. The following theorem relates causality and adjoints.

Theorem 1.8.3: If $W \in \mathcal{J}_+$ is a causal map from Hilbert space into itself, then its adjoint W^* is completely anticipatory.

Proof: If $y \in M_-$, and $x \in M_+$, then $\forall$ such x, y

$$< y, x> = 0 \quad .$$

Now since $W \in \mathcal{J}_+$, $Wx \in M_+$, $\forall\ x \in M_+$

i.e. $$< y, Wx> = 0$$

i.e. $$< W^* y, x > = 0 \qquad \forall\ x \in M_+,\ y \in M_-$$

i.e. $$W^* y \in M_- \qquad \forall\ y \in M_-$$

i.e. $$W^* \in \mathcal{J}_- .$$

Many of the common causal operators used as models in control systems design are non-invertable compact operators. There is an important sub-set of causal operators such that, while the operator G^{-1} may not be bounded, the operator $(\varepsilon I + G)^{-1}$ does exist, and is bounded for all small ε. I.e. $0 < \varepsilon < \varepsilon_{max}$. Following control systems terminology, such an operator is said to have infinite gain margin. This term arises when the operator G can be considered part of a feedback control scheme, as shown in Figure 1.8.1.

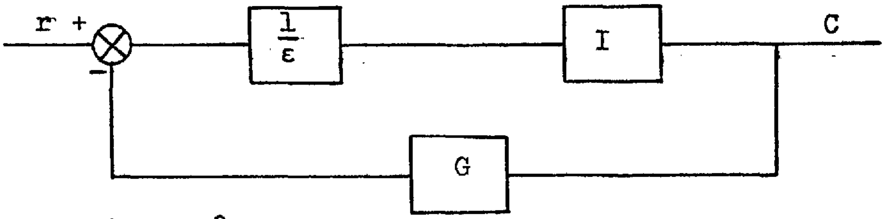


fig. 1.8.1

X

$$C = (\varepsilon I + G)^{-1} . \frac{r}{\varepsilon}$$

The term $\frac{1}{\varepsilon}$ is the "gain" of the closed loop system. The operator G may be unbounded.

Another important concept in control theory is that of minimum phase. For operators specified as rational transfer functions, minimum phase implies no zeros in the (strict) right half complex plane. Now if these transfer functions are inverted, the zeros will become poles. Hence a minimum phase transfer function has an inverse with no poles in the strict right half plane. We generalise these concepts to operators on any space as follows. A causal operator W is said to be minimum phase if there exists a causal operator $W(\varepsilon)$, depending continuously on the parameter ε, such that $W^{-1}(\varepsilon)$ exists, and is bounded for all $0 < \varepsilon < \varepsilon_{max}$, and

$$W \cdot W^{-1}(\varepsilon) \longrightarrow I \qquad \text{as} \qquad \varepsilon \longrightarrow 0$$

where the convergence is with respect to the operator norm. If there exists <u>no</u> causal $W(\varepsilon)$ with this property, the operator is non-minimum phase. This definition is seen to be consistent with the usual definition via frequency response in the case of single-input/single-output time-invariant systems mapping $L_2\,[o, \infty)$. Note that operators with infinite gain margin are always minimum phase.

Let $\quad y(s) = W(s)\,u(s)$.

<u>Example 1.8.1</u>

$$W(s) = \frac{s + a}{(s + b)(s + c)} \qquad\qquad a > 0$$

is minimum phase. For let

$$W(s, \varepsilon) = \varepsilon + W(s)$$

Then $\quad (\varepsilon + W(s))^{-1} = \dfrac{(s + b)(s + c)}{(s + a) + \varepsilon\,(s + b)(s + c)}$

$$= \frac{(s + b)(s + c)}{\varepsilon\, s^2 + s(1 + \varepsilon\,(b + c)) + a + \varepsilon b\, c}$$

which, for ε small enough is always stable.

Obviously $\quad \lim\limits_{\varepsilon \to 0} \; (\varepsilon + W(s))^{-1} W(s) = 1$

Example 1.8.2

$W(s) = \frac{s - a}{s + a}$ is non-minimum phase for $a > 0$. If it were minimum phase, then $\exists\ W^{-1}(s, \varepsilon)$ as above, such that

$$\| W^{-1}(s, \varepsilon)\, W(s) - 1 \| < \delta \quad \text{for } \varepsilon \text{ small enough.}$$

i.e. $$\| \frac{s - a}{s + a} W^{-1}(s,\varepsilon) - 1 \| < \delta$$

i.e. $$\| \frac{s - a}{s + a} W^{-1}(s,\varepsilon) - \frac{s + a}{s - a} \| < \delta$$

But $\frac{s - a}{s + a}$ is all-pass, and so

$$\| \frac{s - a}{s + a} [W^{-1}(s,\varepsilon) - \frac{s + a}{s - a}] \|$$

$$= \| W^{-1}(s,\varepsilon) - \frac{s + a}{s - a} \| < \delta$$

But $\frac{s + a}{s - a}$ is unbounded.

Therefore, for small enough ε, $W^{-1}(s,\varepsilon)$ must be unbounded.

Note that we have used the term unstable in the conventional sense. We shall equate stability with boundedness in this thesis. It is then seen to be a norm dependent property. Consistent with this conventional terminology, an isometric operator that is causal and non-minimum phase is called all-pass.

1.9 Contraction and Spectral Theory

An operator A in a metric space X is said to be a contraction if for all $x, y \in X$

$$\rho(A x, A y) < K \rho(x, y) , \qquad 1.9.1$$

where $K < 1$ is a positive constant.

A fixed point of an operator A is x such that

$$A x = x . \qquad 1.9.2$$

Now the contraction operators are important due to the theorem of Banach, which states:

Theorem 1.9.1: If the contraction operator A maps a complete metric space X onto itself, then we have a unique fixed point and this point can be obtained by the method of successive approximation for any initial point $x_o \in X$.

Proof: [V 1; p.115]

This theorem finds its main application in the theory of the solution of operator equations of the form

$$\lambda y - A y = b \qquad 1.9.3$$

or

$$y = \frac{1}{\lambda} A y + \frac{b}{\lambda} . \qquad 1.9.4$$

Then it is possible to find conditions on λ under which the operator on the left side of 1.9.4 represents a contraction. An operator A

may not be a contraction, but a finite power A^N may be. In this case, theorem 1.9.1 still holds for A. Using this fact, it can be shown that the causal operator given by

$$y(t) = \int_0^t W(t,\tau)\, u(\tau)\, d\tau \qquad 1.9.5$$

represents a contraction, [T 1; p.168-9], for y, u on finite intervals. In fact, with this Volterra operator, equation 1;9.4 can be solved by successive approximation for any non-zero λ.

The solvability of equation 1.9.3 is very important in control studies, and the parameter λ plays an important role. Spectral theory is concerned with investigating the set of (complex) scalars λ and vectors $x \in X$, such that

$$A\,x = \lambda\,x\,. \qquad 1.9.6$$

Such λ are called eigenvalues, and the corresponding x are eigenvectors. The theory of spectral analysis is well known and we shall merely state a few results which we used in the sequel. The set of eigenvalues of A is called the spectrum of A, denoted $\sigma(A)$, and its complement is the resolvent set. The operator

$$R(\lambda, A) = (\lambda I - A)^{-1} \qquad 1.9.7$$

defined on the resolvent set, is called the resolvent of A. It is known [DS 1; p.568] that

$$\sigma(A) = \sigma(A^*) \qquad 1.9.8$$

Some important results that will be used are as follows:

Theorem 1.9.2: If T is a compact operator, its spectrum is at most denumerable and has no point of accumulation in the complex plane except possibly at $\lambda = 0$.

Proof: [DS 1; p.579].

Theorem 1.9.3: If T is self-adjoint, and λ is an eigenvalue of T, then λ is real, and eigenvectors corresponding to distinct eigenvalues are orthogonal.

Proof: [T 1; p.324].

Corollary; If A is self-adjoint on a Hilbert space, with eigenvalues λ,

$$\|A\| = |\lambda|_{\max}$$

and if A^{-1} exists and is bounded

$$\|A^{-1}\| = \frac{1}{|\lambda|_{\min}}$$

Proof: [T 1; p.325].

Theorem 1.9.4: If T is a compact mapping from a Banach space S onto itself, and $\lambda \neq 0$, the equation $(\lambda I - T)x = y$ has a unique solution $\forall y \in X$, if and only if the equation $(\lambda I - T)x = 0$ has no solution other than $x = 0$.

Proof: [DS 1; p.515].

By the contraction mapping principle, it has been shown that, for all λ, the Volterra integral equation over a finite interval

$$\lambda x - W x = y$$

has a unique solution for all y, and hence the only eigenvalue of W is zero, by theorem 1.9.4. The same reasoning also applies to causal weighting sequences W_{ij}, where $W_i = 0$ for all i. The difficulty in the argument for infinite intervals is that $(\lambda I - W)^{-1}$ may not be bounded for some $|\lambda| < |\lambda|_{max}$.

1.10 Semi-groups of Operators and State Space

The concept of a semi-group of operators was introduced in connection with stationarity. In this section, more general semi-groups of operators are discussed, and the concept of state is introduced. We follow closely the notation&terminology of Dunford and Schwartz, Vol. 1, Ch. VIII. A one-parameter semi-group is called uniformly continuous if the operator $f : t \longrightarrow T(t)$ is continuous with respect to the norm of T. [DS1; p.614].

Theorem 1.10.1: Let $T(t)$ be a uniformly continuous semi-group. Then there exists a bounded operator A, such that $T(t) = e^{At}$ for $t \geq 0$. The operator A is given by the formula $A = \lim_{h \to o} (T(h) - I)/h$.

For the real part of λ sufficiently large, the resolvent of A can be expressed in terms of the Laplace transform of the semi-group, by the formula

$$R(\lambda, A) = (\lambda I - A)^{-1} = \int_{o}^{\infty} e^{-\lambda t} T(t)\, dt \; . \qquad 1.10.1$$

Proof: [DS 1; p.615].

A is called the infinitesimal generator of the semi-group, and we may write

$$\frac{dT}{dt} = A\, T(t) \; . \qquad 1.10.2$$

A semi-group may only be strongly continuous, rather than uniformly continuous [DS 1; p.614]. If it is not uniformly continuous, it will have an infinitesimal generator which is unbounded, since uniform continuity is necessary and sufficient for boundedness of A.

Example 1.10.1: The shift operators on $L_2 [o, \infty)$ constitute a strongly continuous semi-group, which is not uniformly continuous. The infinitesimal generator is then the (unbounded) operation of differentiation $\frac{d}{dt}$. This corresponds to the Laplace transform of the delay operator e^{sT}.

Example 1.10.2: The semi-group of operators e^{At} where A is an $n \times n$ matrix is uniformly continuous with infinitesimal generator A, operating on n-dimensional euclidean space.

Let $T(t)$ be a map from the space X onto itself. Then $x(t) \in X$ for each fixed t, where

$$x(t) = T(t)\, x_o \quad , \quad x_o \in X \quad . \qquad 1.10.3$$

$$\frac{dT(t)}{dt} = A\, T(t) \quad .$$

Therefore

$$\frac{dx}{dt} = A\, x \quad . \qquad 1.10.4$$

This differential equation represents the autonomous evolution of a trajectory of vectors from an initial vector x_o. For control systems studies, we may wish to add forcing terms, and observation terms to result in

$$\left.\begin{aligned} \frac{dx}{dt} &= A\, x + B\, u \\ y &= C\, x + D\, u \,. \end{aligned}\right\} \qquad 1.10.5$$

The y variable is the output variable at time t, and the u variable is the control. A, B and C are linear operators in appropriate vector spaces. This description 1.10.5 is called the state space description of linear systems. For added generality, it may be assumed that A, B and C are functions of time. The variable x is called the state of the system at time t. It is easy to rewrite 1.10.5 in terms of the explicit notation for linear systems.

$$y(t) = C(t)\,\Phi(t, t_o)\,x_o + C(t)\int_o^t \Phi(t,\tau)\,B\,u(\tau)\,d\tau$$

$$= y_o \qquad + W u \qquad 1.10.6$$

where Φ is the transition operator of the system.

The dimension of the state-space is called the dimension of the system. If it is finite dimensional, then 1.10.5 is a set of simultaneous ordinary differential equations representing lumped systems as discussed in section 1.1. However, 1.10.5 is more general than this if the system is infinite dimensional.

Example 1.10.3: [DS 1; p.614]. An example of an autonomous system where the state is infinite dimensional is the system described by

$$\frac{\partial x}{\partial t} = \frac{\partial^2 x}{\partial s^2} + h(s)\,x + \int_o^\infty K(s, u)\,x\,(u, t)\,du \qquad 1.10.7$$

$$x(s, o) = x_o(s)$$

where $t \in [o, \infty)$, $s \in [o, \infty)$. This is of the form

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

where A is the unbounded operator on the left of equation 1.10.7.

There are equivalent forms for discrete time systems where the semi-group is no longer continuous. We then use difference equations, and the concept of semi-group generator no longer becomes necessary.

It is not at all clear that all linear system operators can be represented in state-space form, though, as will be pointed out in the next chapter, this representation has interesting theoretical advantages. The point of view adopted in this thesis is that linear operators will be investigated independently of semi-group theory and state-space, and then comparisons will be made between these results and known state-space results. In particular, we shall try and use the insight that the state-space description gives to design control systems for more general models.

CHAPTER 2

OPTIMAL CONTROL

2.1 Performance of Control Systems

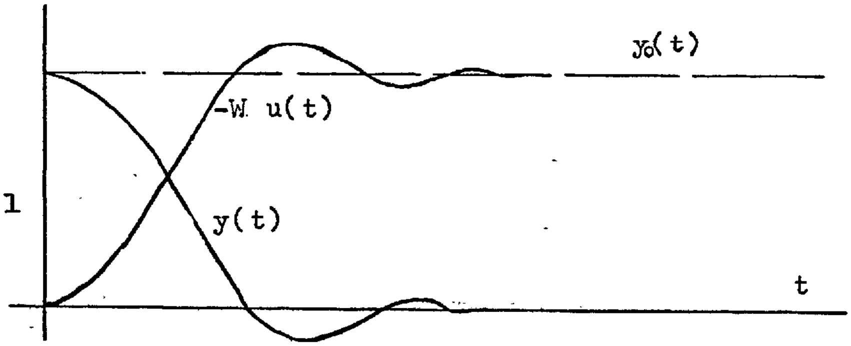
For the system represented by

$$y = y_o + W u \qquad 2.1.1$$

the general control problem is to choose u from an allowable set so that the resulting y is good in some sense. It may be quite difficult to attach a precise meaning to "good" in an engineering problem. In this section, some engineering performance criteria are examined. Firstly, we restrict our discussion to continuous single input single output (SI/SO) causal systems. That is, y and u lie in some appropriate space of scalar time functions defined on an interval $[0, T]$, or $[0, \infty)$. The signal $y_o(t)$ may either represent disturbances due to initial conditions of the system, or some external influence such as change of desired set point, or a superposition of both these effects.

Engineering requirements usually dictate that the difference between y_o and $(- W u)$, i.e. y, should be small in some sense. The signal y usually represents the error between the actual and desired performance. On the other hand, it may only be possible to make y small by making the control u large, which can be undesirable, and an engineering compromise between smallness of y and largeness of u has to be made. A large class of simple systems can be specified in terms of their response to a step signal. If $y_0(t) = 1, \; t > 0$, then a typical desirable response of $-Wu$ is as shown in Figure 2.1.1.

fig. 2.1.1



We may require either zero or very small steady state error , a fairly quick response, but with not too much overshoot, and also a quick decay of any oscillations. If this kind of response can be obtained without too much control effort, the system designer is usually content. Of course, other criteria are also specified, such as noise rejection, velocity lag, sensitivity, etc.

It is difficult to sum up all these criteria in terms of one number. However, to compare various alternatives, and for automatic design, such a cost functional is needed. Norms are a convenient scalar measure of the size of a signal, though they are inadequate to express properties of signal shape. However, it is possible to use weighted norms, by emphasising different instants of time. A typical cost function becomes

$$J = f(\|y\|_{w_1(t)} , \|u\|_{w_2(t)}) \qquad 2.1.2$$

A useful and mathematically convenient norm is the Hilbert space norm. The cost is taken to be

$$J = \langle y, Q\, y \rangle + \langle u, R\, u \rangle \qquad 2.1.3$$

where Q and R are self-adjoint positive operators. In L_2, Q and R can be positive scalar time functions, and then

$$J = \int_0^T (y^2(t)\, q(t) + u^2(t)\, r(t))\, dt \; . \qquad 2.1.4$$

The relative weightings Q and R depend on the designer. The smaller that R becomes, relative to Q, the larger the control allowed for a fixed control cost, and hence it is to be expected that the output response should be better. A system with a good value of quadratic performance criterion should tend to have a response which is neither overdamped nor over oscillatory.

This criterion, however, does have some disadvantages. On the one hand, it can allow a finite cost to a system with infinitely fast oscillations confined in an exponential envelope as in Figure 2.1.2.

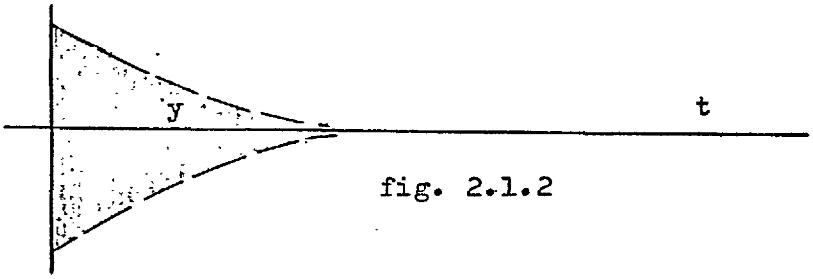


fig. 2.1.2

On the other hand, it may give an infinite cost to a system with non-zero steady state offset, which is not necessarily bad. This latter difficulty can be removed by the choice of a decaying $q(t)$, $r(t)$, but it is often convenient to choose q and r constant numbers.

With multivariable systems, the performance criteria become even more difficult to specify. Cross-coupling and interaction between outputs may or may not be considered good. Conflicting requirements on the various outputs make design very difficult, and in fact there are no simple design methods for multivariable systems which enable the designer to obtain an intuitive feel for the problem, as there are for SI/SO systems. In particular, the Q and R operators become matrix functions of time, whose elements are at the designer's choice.

Performance criteria of the type 2.1.3 will be used primarily throughout this thesis, and u, y will be considered elements of Hilbert spaces $\mathcal{H}u$, $\mathcal{H}y$. Having chosen such a cost function, it is natural to ask the question: what $u \in \mathcal{H}_u$ minimises J ? This is answered in the next section.

2.2 Conditions of Optimality

The work of Hsieh [H 4] is of fundamental importance to the next few sections. However, the derivations and proofs presented here differ from Hsieh's work. The theorem statements are more explicit than Hsieh's, and the proofs are original, and more direct.

Consider the system

$$y = y_o + W u \qquad 2.2.1$$

with $u \in \mathcal{H}_u$, $y \in \mathcal{H}_y$, where $\mathcal{H}_u$ and $\mathcal{H}_y$ are Hilbert spaces. We propose the cost function

$$J = < y, Q y > + < u, R u > \qquad 2.2.2$$

where the operators Q and R are self-adjoint mappings of $\mathcal{H}_y$ and $\mathcal{H}_u$ respectively. As distinct from the approach of Hsieh, we first assume that there exists a $\hat{u} \in \mathcal{H}_u$, which produces a $\hat{y} \in \mathcal{H}_y$ according to equation 2.2.2, such that there is no other u which produces a lower cost.

Theorem 2.2.1: If the operators R and Q are positive semi-definite and bounded, anu W is bounded, then a necessary and sufficient condition for optimality of the control $\hat{u} \in \mathcal{H}_u$, is that the vector $\hat{g} \in \mathcal{H}_u$, given by

$$\hat{g} = R\hat{u} + W^* Q \hat{y} \qquad 2.2.3$$

is zero.

<u>Proof</u>: For any $u \in \mathcal{H}_u$,

$$J = \langle y, Q y \rangle + \langle u, R u \rangle$$

$$= \langle \hat{y} + \Delta y, Q(\hat{y} + \Delta y) \rangle + \langle \hat{u} + \Delta u, R(\hat{u} + \Delta u) \rangle$$

$$= \langle \hat{y}, Q \hat{y} \rangle + \langle \hat{u}, R \hat{u} \rangle + 2\langle \Delta y, Q \hat{y} \rangle + 2\langle \Delta u, R \hat{u} \rangle$$

$$+ \langle \Delta y, Q \Delta y \rangle + \langle \Delta u, R \Delta u \rangle$$

$$= \hat{J} + 2\langle \Delta y, Q \hat{y} \rangle + 2\langle \Delta u, R \hat{u} \rangle + \langle \Delta y, Q \Delta y \rangle + \langle \Delta u, R \Delta u \rangle .$$

But $\Delta y = W \Delta u$. 2.2.4

So $J - \hat{J} = 2\langle W\Delta u, Q \hat{y} \rangle + 2\langle \Delta u, R \hat{u} \rangle + \langle W \Delta u, Q W \Delta u \rangle$

$$+ \langle \Delta u, R \Delta u \rangle$$

$$= 2\langle \Delta u, R \hat{u} + W^* Q \hat{y} \rangle + \langle \Delta u, (W^* Q W + R) \Delta u \rangle . \quad 2.2.5$$

Certainly, if $\hat{g} = 0$, then

$$J - \hat{J} = \langle \Delta u, (W^* Q W + R) \Delta u \rangle \quad 2.2.6$$

$$\geqq 0 \qquad \forall \Delta u \quad ,$$

since $R + W^* Q W$ is positive semi-definite.

However, $\hat{g} = 0$ is also necessary, for, if $\hat{g} \neq 0$, then choose $\Delta u = - \varepsilon \hat{g}$, which we can do, since Δu is arbitrary, and $\varepsilon > 0$ is a scalar. Note that $\hat{g} \in \mathcal{H}_u$, since all the operators in 2.2.3 are bounded, and $\hat{u}$, $\hat{y}$ are Hilbert space elements. Then consider two cases.

(i) $< \hat{g}, (W^{*} Q W + R) \hat{g} > = 0$, but $\hat{g} \neq 0$.

Then
$$J - \hat{J} = - \varepsilon < \hat{g}, \hat{g} >$$
$$< 0 \quad \forall \; \varepsilon > 0$$

which is a contradiction to $\hat{J}$ being the optimal cost.

(ii) $< \hat{g}, (W^{*} Q W + R) \hat{g} > \neq 0$, $\hat{g} \neq 0$.

Then choose
$$\varepsilon = \frac{< \hat{g}, \hat{g} >}{< \hat{g}, (W^{*} Q W + R) \hat{g} >} > 0, \text{ finite}. \qquad 2.2.7$$

$$J - \hat{J} = - 2 \varepsilon < \hat{g}, \hat{g} > + \varepsilon^{2} < \hat{g}, (W^{*} Q W + R) \hat{g} >$$
$$= - \varepsilon < \hat{g}, \hat{g} > \left(2 - \frac{\varepsilon < \hat{g}, (W^{*} Q W + R) \hat{g} >}{< \hat{g}, \hat{g} >}\right),$$
$$= - \varepsilon < \hat{g}, \hat{g} >$$
$$< 0$$

which is again a contradiction to optimality.

Therefore $\hat{g} = 0$ is a necessary and sufficient condition for optimality.

Two questions have now to be answered. Is there any $\hat{u} \in \mathcal{H}_{u}$, which satisfies the above conditions, and if so, is it unique?

Theorem 2.2.2: Under the conditions of Theorem 2.2.1, a unique $\hat{u}$ exists if $R + W^* Q W$ has a bounded inverse and $y_o \in \mathcal{H}_y$.

Proof: $W^* Q$ is a bounded mapping from $\mathcal{H}_y$ to $\mathcal{H}_u$.

So
$$f = - W^* Q y_o$$
$$\in \mathcal{H}_u .$$

Consider the unique vector

$$\hat{u} = - (R + W^* Q W)^{-1} W^* Q y_o \qquad 2.2.8$$
$$= (R + W^* Q W)^{-1} f$$
$$\in \mathcal{H}_u .$$

But
$$(R + W^* Q W) \hat{u} + W^* Q y_o = 0 .$$
$$R \hat{u} + W^* Q (y_o + W \hat{u}) = 0 .$$

Let
$$\hat{y} \in \mathcal{H}_y$$
$$= y_o + W \hat{u} .$$

Then
$$R \hat{u} + W^* Q \hat{y} = 0 .$$

i.e. $\exists$ $\hat{u} \in \mathcal{H}_u$ satisfying the necessary and sufficient conditions for optimality.

In Theorem 2.2.2, $(R + W^* Q W)^{-1}$ bounded is only a sufficient condition for existence of $\hat{u}$, but not necessary. If the vector $W^* Q y_o \in \mathcal{R}(R + W^* Q W)$ then $\hat{u}$ still exists in $\mathcal{H}_u$. However, $\hat{u}$ may no longer be unique.

Theorem 2.2.3: The optimal cost $\hat{J}$ is given by

$$\hat{J} = \langle y_o, Q \hat{y} \rangle \qquad 2.2.9$$

Proof: $\hat{J} = \langle \hat{y}, Q \hat{y} \rangle + \langle \hat{u}, R \hat{u} \rangle$

$$= \langle y_o + W \hat{u}, Q \hat{y} \rangle + \langle \hat{u}, R \hat{u} \rangle$$

$$= \langle y_o, Q \hat{y} \rangle + \langle W \hat{u}, Q \hat{y} \rangle + \langle \hat{u}, R \hat{u} \rangle$$

$$= \langle y_o, Q \hat{y} \rangle + \langle \hat{u}, W^* Q \hat{y} + R \hat{u} \rangle$$

$$= \langle y_o, Q \hat{y} \rangle$$

In the preceding work, it has been necessary to assume $W: \mathcal{H}_u \rightarrow \mathcal{H}_y$ is a bounded operator, and $y_o \in \mathcal{H}_y$. This may exclude initially unstable systems.

However, the vector $y \in \mathcal{H}_y$ may be achieved by $y_o + W u \in \mathcal{H}_y$ without either y_o or $W u \in \mathcal{H}_y$, if the sum is defined appropriately. Hence, with W unbounded and a y_o from a particular class, it may be possible to find a $u \in \mathcal{H}_u$ which can stabilise the system. The difficulty in the analysis of unbounded systems is that unboundedness and discontinuity are synonymous for linear operators. This means that if u gives a finite cost, $u + \Delta u$ may give an infinite cost, no matter how small $\| \Delta u \|$ is chosen. This topic is taken up in Section 2.7.

The classical method of solution of functional equations of the type

$$(R + W^* Q W) \hat{u} = - W^* Q y_o \qquad 2.2.10$$

is in terms of eigenfunctions and eigenvalues of the operator $W^* Q W$ or associated operators. In fact, equation 2.2.10 is first converted to

$$(I + R^{-\frac{1}{2}} W^* Q W R^{-\frac{1}{2}}) v = - R^{-\frac{1}{2}} W^* Q y_o \qquad 2.2.11$$

where it is assumed that $R^{-\frac{1}{2}}$ exists, bounded, and is self-adjoint

and $$v = R^{+\frac{1}{2}} \hat{u} . \qquad 2.2.12$$

Then if W is compact, so is $R^{-\frac{1}{2}} W^* Q W R^{-\frac{1}{2}}$ (being the product of bounded and compact operators), and hence it has a denumerable spectrum with eigenvalues λ_i, and associated eigenvectors φ_i. However, $R^{-\frac{1}{2}} W^* Q W R^{-\frac{1}{2}}$ is self-adjoint, so the λ_i are real, and the φ_i are orthogonal. Moreover, the only accumulation point of the λ_i is zero.

Following the procedure of Hsieh [H 4; p.134] or Riesz-Nagy [RN 1; p.235] we obtain

$$\hat{u} = - R^{-\frac{1}{2}} (W^* Q y_o + \sum_i \frac{< Q y_o, W R^{-\frac{1}{2}} \varphi_i >}{1 + \lambda_i} \varphi_i) \qquad 2.2.13$$

where the φ_i are assumed normalised.

This approach is not at all convenient for practical computation. In fact, it is just a method for writing

$$\hat{u} = - (R + W^* Q W)^{-1} W^* Q y_o$$

in terms of an eigenfunction expansion, where calculation may be far more work than evaluating the inverse of $R + W^* Q W$ explicitly. The next few sections develop feasible algorithms for computation of optimal controls. They represent a considerable extension of Hsieh's work.

2.3 Gradient Algorithms

While the eigenvalue-eigenvector solution presented in the last section gives an explicit solution to the optimal control problem, it is often quicker to find the optimal control vector than to find eigenfunctions, without explicit inversion of the operator $R + W^* Q W$. Among the methods that exist for finding optimal controls are the descent methods which directly minimise the cost function. Descent methods are iterative methods, and find a sequence of control vectors which directly minimise the cost function, such that the cost is monotonically reduced at each iteration. They have the advantage that the algorithm may be terminated after a finite number of iterations, to provide a sub-optimal control which may give a satisfactory cost. The basis of these methods is the following:

$$J = \langle y, Q\, y \rangle + \langle u, R\, u \rangle$$

If u is changed to $u + \Delta u$, then, similar to the derivation of section 2.2

$$\Delta J = 2\langle \Delta u, R\, u + W^{*} Q\, y \rangle + \langle \Delta u, (R + W^{*} Q\, W)\Delta u \rangle$$

It has been shown that if $g = R\, u + W^{*} Q\, y$ is non-zero, it is possible to find an $\varepsilon > 0$, such that $\Delta u = -\varepsilon\, g$ will allow a decrease in cost. This procedure gives rise to the method of steepest descent, but there are other ways of choosing Δu which also decrease the cost at each iteration. All first order descent algorithms can be put into a standard form:

<u>Algorithm 2.3.1:</u>

1. Choose a nominal control u_o (perhaps zero).

2. Calculate y_i from the control u_i via the dynamic equations.

3. Calculate the gradient g_i.

4. Test stopping procedure for convergence.

5. From g_i and any other previous information, calculate a search direction p_i.

6. Let $u_{i+1} = u_i + \alpha_i p_i$, where α_i is a scalar chosen to give a decrease in cost.

7. Go back to 2.

In step 6, it is often most efficient computationally to choose the scalar α_i to maximise the decrease in cost in the direction p_i.

Theorem 2.3.1: The optimal α_i, in direction p_i, is given by

$$\alpha_i = - \frac{< p_i, R u_i> + <W p_i, Q y_i>}{< p_i, R p_i> + <W p_i, QWp_i>} . \qquad 2.3.1$$

Proof: $\Delta J = 2<\Delta u, g_i> + <\Delta u, (R + W^* Q W) \Delta u> .$ 2.3.2

But $\Delta u = \alpha_i p_i$.

So $\Delta J = 2 \alpha_i < p_i, g_i > + \alpha_i^2 <p_i, (R + W^* Q W) p_i> .$

But ΔJ is a differentiable function of α, with positive (constant) second derivative. So a unique minimum is given by

$$\frac{\partial \Delta J}{\partial \alpha_i} = 0 .$$

i.e. $2 <p_i, g_i> + 2 \alpha_i < p_i, (R + W^* Q W) p_i > = 0$.

i.e. $$\alpha_i = - \frac{<p_i, g_i>}{<p_i, (R + W^* Q W) p_i>}$$

$$= - \frac{<p_i, R u_i> + <W p_i, Q y_i>}{<p_i, R u_i> + <W p_i, QWp_i>} .$$

The three main gradient procedures used in practice are

(a) Steepest descent

(b) Partan

(c) Conjugate gradient (Fletcher and Reeve's Method [FR1])

The main difference between these methods occurs in step 5 of the general algorithm 2.3.1.

Algorithm 2.3.1a: Steepest Descent.

Step 5 becomes $p_i = g_i$.

Hsieh [H4; p.141] has shown that the steepest descent method produces a sequence of costs which converge to the optimal cost, and also the sequence of control vectors converge strongly to the optimal control vector. Steepest descent can have rather poor convergence properties. Kantorovich has derived a bound for the rate of convergence of this method. Specifically, if

$$m <x, x> \leqq <x, (R + W^* Q W) x> \leqq M <x, x> \qquad 2.3.3$$

then $$\| u_n - \hat{u} \| \leqq \frac{1}{m} \left(\frac{M - m}{M + m} \right)^n (J(u_o) - J(u))^{\frac{1}{2}} \qquad 2.3.4$$

$$= K \left(\frac{1 - \frac{m}{M}}{1 + \frac{m}{M}} \right)^n . \qquad 2.3.5$$

The smaller $\frac{m}{M}$, the inverse of the condition number, the poorer the convergence. This poor convergence is interpreted geometrically as the "ridge phenomenon", where the algorithm takes small steps backward and forward across a ridge, as illustrated for a 2-dimensional Euclidean space in Figure 2.3.1.

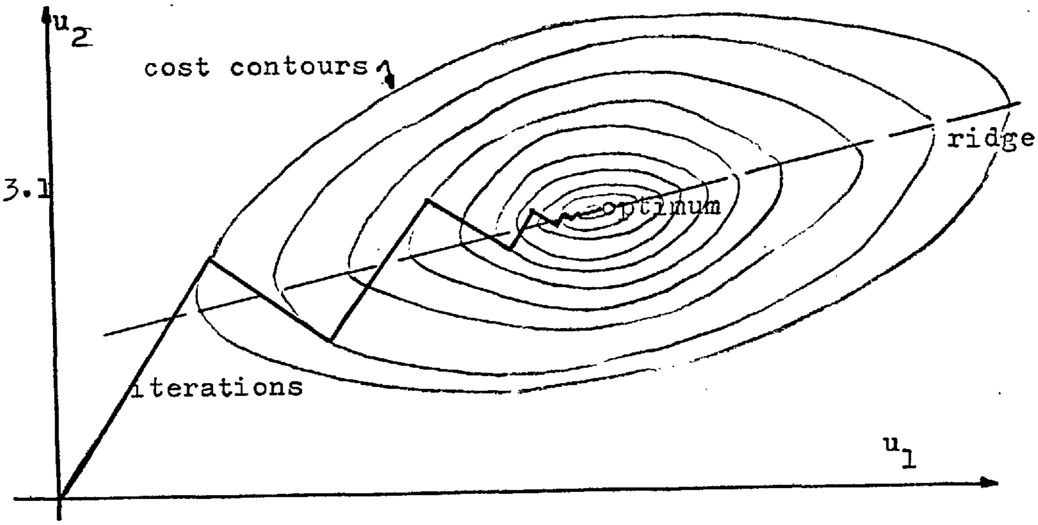


fig. 2.3.1

The ratio $\frac{m}{M}$ is a measure of the difficulty of a problem. Hsieh introduces a modified steepest descent algorithm which tries to speed up convergence by performing many simple steepest descent iterations in one modified iteration. It is mainly the poor convergence properties of the steepest descent type algorithms which prompts research into other methods of choosing a search direction. While it is known that, for steepest descent, if there is no convergence at the first iteration, there can be no finite convergence, there do exist methods which have finite convergence in finite dimensional Euclidean space, for the quadratic problems considered here. The first of these is Partan, which modifies the general algorithm as follows.

Algorithm 2.3.1b: Partan.

Renumber the iterations 0, 2, 3, ... (i.e. omit 1). Then Step 5 becomes

5 (i) At a point of even subscript $p_{2m} = g_{2m}$.

5 (ii) At a point of odd subscript $p_{2m+1} = u_{2m+1} - u_{2m-2}$. 2.3.7

The heuristic basis for this algorithm is that it tends to line up the search directions along a ridge, rather than allow them to zig-zag across it. In an n-dimensional space, this method will converge to the exact optimum in at most $2n$ iterations.

The most efficient first order algorithm is the conjugate gradient method of Fletcher and Reeves.

Algorithm 2.3.1c: Conjugate Gradient.

Step 5 becomes:

Calculate g_i .

$$\text{Calculate} \quad \beta_i = \frac{<g_i, g_i>}{<g_{i-1}, g_{i-1}>} \quad . \qquad 2.3.8$$

$$p_i = - g_i + \beta_i \, p_{i-1} \quad . \qquad 2.3.9$$

A basic reference for the properties of this method in Hilbert space is the paper by Antosiewicz and Rheinboldt [AR 1] Lasdon, Mitter and Waren [LMW 1] also prove some results of interest to control systems. Both papers state the following properties.

(1) The set p_i are orthogonal with respect to the operator $R + W^* Q W$.

$$\text{i.e.} \quad < p_i, (R + W^* Q W) p_j > = 0 , \qquad i \neq j \quad . \qquad 2.3.10$$

(2) At the i:th iteration, J is minimised over the set u, such that

$$u = u_o + \sum_{j=1}^{i} \gamma_j p_j \quad .$$

This set is denoted by B_i.

(3) The norm of the error $\| u_i - \hat{u} \|$ and the cost J are decreased at each step.

(4) If $\mathcal{H} = \bigcup_i B_i$, then the sequence of u_i converge

strongly to the optimal $\hat{u}$, and the sequence of cost functions J_i converge to the optimal $\hat{J}$, if $u_o \in \mathcal{H}_u$.

The property (2) makes this algorithm optimal, and implies finite convergence for finite dimensional problems. In fact, the conjugate property (1) implies (2) as follows:

$$g_{i+1} = g_i + \Delta g_{i+1} \qquad 2.3.11$$

and $$g_{i+1} = A \Delta u_i \qquad 2.3.12$$

where $$A = R + W^* Q W \qquad 2.3.13$$

$$g_i = A u_i + b \qquad 2.3.14$$

and $$g_{i+1} = A u_{i+1} + b \quad .$$

But $$\Delta u_i = \alpha_i p_i \quad . \qquad 2.3.15$$

So $$g_{i+1} = g_i + \alpha_i A p_i \quad . \qquad 2.3.16$$

Now, for property (2) to hold, g_{i+1} is required to be orthogonal to <u>all</u> previous search directions p_i

i.e. $$< p_j, g_{i+1} > = 0 \qquad j = 1 \ldots i \qquad 2.3.17$$

i.e. $$< p_j, g_i > + \alpha_i < p_i, A p_i > = 0 \quad . \qquad 2.3.18$$

Now this property holds up to iteration i, for $j = 1 \ldots i-1$, if $< p_i, A p_i > = 0$. It is also true for $i = 1$. But by theorem

2.3.1, α_i is chosen so that 2.3.18 holds for $j = i$. Hence by induction (2) is true $\forall$ i.

The choice of β_i can be shown to preserve the conjugacy of the search directions.

Lasdon et al. [LMW 1] show that this method is always better than the steepest descent method. Antosiewicz and Rheinboldt [AR 1] give similar geometric convergence rates to equation 2.3.4, which depend on the convergence factor

$$\left(\frac{M - m}{M} \right) = \left(1 - \frac{m}{M}\right)$$

where m, M are defined in equation 2.3.3. This algorithm takes full account of the eccentricity of the cost contours, and theoretically avoids the ridge behaviour. However, numerical inaccuracies may upset these properties as is pointed out in the next section, where computed examples of control trajectory design are presented.

2.4 Computational Examples

This section presents some computed optimal control examples to illustrate the properties of the gradient algorithms, primarily the conjugate gradient algorithm. Computational problems are also discussed, including the truncation of the time axis, and numerical errors with ill-conditioned problems.

In computation of trajectories, only finite intervals of time can be considered. Hence, we are required to approximate an infinite interval optimisation problem by a finite one; i.e. the time axis is truncated after a time T. Then, if the computed trajectories of input and output converge to zero in the time interval $[0, T]$, and the assumption is made that they remain at zero from $[T, \infty)$, then these trajectories approximate the infinite time optimal trajectories. This assumption is quite difficult to justify in general. In a causal system, the output trajectory at time t only depends on the past, and so the important parts of y_o and W are the relevant restrictions to the truncation interval. The optimal control is determined by the future behaviour of the optimal output, but if this decays to zero, then the far future will have negligible effect on the present. This restriction of y_o and W to the smaller space may seem rather drastic truncation, especially when y_o and W are unbounded on the infinite interval, but bounded on a finite interval, as is the usual case for unstable open-loop systems. The justification of the method is a problem for future research, though some of the difficulties which arise in practice are discussed in the sequel.

For multivariable control problems the expressions for the dynamics, gradient and cost are as follows.

(1) Sampled data:

$$y_k = y_{o_k} + \sum_{j=0}^{k} W_{k,j} u_j \qquad 2.4.1$$

$$g_k = R_k u_k + \sum_{j=k}^{N} W^T_{j,k} Q_j y_j \qquad 2.4.2$$

$$J = \sum_{k=0}^{N} y_k^T Q_k y_k + u_k^T R_k u_k . \qquad 2.4.3$$

(2) Continuous time:

$$y(t) = y_o(t) + \int_0^t W(t, \tau) u(\tau) d\tau \qquad 2.4.4$$

$$g(t) = R(t) u(t) + \int_t^T W^T(\tau, t) Q(\tau) y(\tau) d\tau \qquad 2.4.5$$

$$J = \int_0^T (y^T(t) Q(t) y(t) + u^T(t) R(t) u(t)) dt . \qquad 2.4.6$$

If $R^{-1}(t)$ is an absolutely continuous function of time, then without loss of generality so is the optimal $u(t)$. For, from equation 2.4.5,

$$u(t) = - R^{-1}(t) . \int_t^T W^T(\tau, t) Q(\tau) y(\tau) d\tau$$

the product of two absolutely continuous functions.

Example 2.4.1: Sampled data.

$$y_{k+1} = y_k + u_k$$

where $$y_o = 2$$

and $$J = \sum_{k=1}^{N} y_k^2 + 0.75 u_k^2 .$$

This system fits neatly into the state-space formulation to be discussed in Section 2.8. If the summation in J is taken over zero to infinity, then the optimal solutions are given by

$$u_k = - 2/3 y_k$$

and $$y_{k+1} = 1/3 y_k$$

with $$J = 6 .$$

Our method sets

$$y_k = y_{o_k} + \sum_{j=o}^{k} W_{k-j} u_j$$

where $$y_{o_k} = 2 \qquad \forall k$$

and $$W_k = 1 \qquad \forall k \neq 0$$

$$= 0 \qquad \text{for } k = 0$$

and $N = 19$ is supposed large enough for the trajectories to converge to zero. In fact, $N = 6$ would have been sufficient. The three algorithms 2.3.1a-c were used to solve this problem, and the relative performance is shown in Figure 2.4.1. The same nominal control of zero was used in each case, and the time per iteration was substantially the same in the three cases.

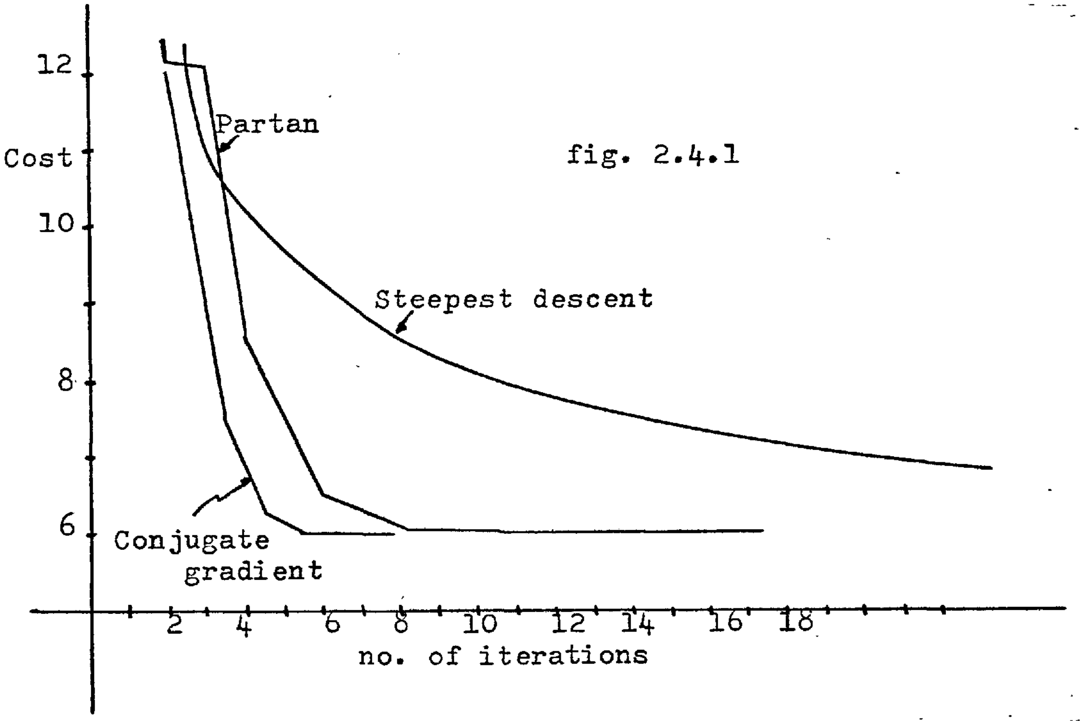


fig. 2.4.1

The convergence bounds of the type 2.3.4 may be calculated. We require m, M, where

$$m \langle x, x \rangle \leqq \langle x, (R + W^* Q W) x \rangle \leqq M \langle x, x \rangle \qquad 2.4.7$$

For SI/SO sampled data systems which are time invariant, the operator W can be represented by a matrix

$$W = \begin{pmatrix} W_o & & & & \bigcirc \\ W_1 & W_o & & & \\ W_2 & W_1 & W_o & & \\ W_3 & W_2 & W_1 & W_o & \\ \vdots & & & & \ddots \end{pmatrix} \qquad 2.4.8$$

If the time axis is truncated, this matrix is finite, and W^* is represented by the transposed matrix. If $W_o = 0$, the matrix is singular, and so is $W^* W$. Hence, for example,

$$m \geq r = 0.75 .$$

Also $$W^* W = \begin{pmatrix} 19 & 18 & 17 & .. & .. & 0 \\ 18 & 18 & 17 & .. & .. & 0 \\ 17 & 17 & 17 & .. & .. & 0 \\ 16 & .. & .. & & & \\ \vdots & & & & & \end{pmatrix}$$

and by Gershgorin's theorem [W 1], the maximum eigenvalue lies in a disc centre 19 and radius 18 x 9

i.e. $$\|W^* W\| \leq 181$$

So $$\frac{M - m}{M + m} \sim 1 - \frac{1.5}{181.75}$$

$$= 1 - 0.0083$$

By equation 2.3.4, the steepest descent method would give three figure convergence in less than n iterations, where

$$\| u - \hat{u} \| = 0.001$$

$$= \left(\frac{M - m}{M + m} \right)^n \frac{\sqrt{3}}{0.75}$$

$$\text{as} \qquad n \sim 930 .$$

This bound is rather crude, but the example serves to show the very large range of eigenvalues that occur in control problems of this sort. It is for precisely these ill-conditioned problems that the conjugate gradient method is best suited. In fact, it is known that the conjugate gradient method converges in a finite number of iterations in the finite dimensional case, and at a rate faster than any geometric series for infinite dimensional problems [H 3], [SL 1]. However, numerical inaccuracies can upset these properties for very small condition numbers, and it is still important to keep these as small as possible. Note that $\| W^{*} W \|$ depends to a high degree on the total costing interval. For continuous time systems (SI/SO), if

$$z(t) = \int_{0}^{t} W(t, \tau)\, u(\tau)\, d\tau \qquad 2.4.9$$

$$|z(t)|^2 \leqq \int_{0}^{t} |W(t, \tau)|^2 \, d\tau \; . \int_{0}^{t} |u(\tau)|^2 \, d\tau \qquad 2.4.10$$

i.e. $$\int_0^T |z(t)|^2 \, dt \leq \int_0^T \left(\int_0^t |W(t,\tau)|^2 \, d\tau \right) dt \cdot \int_0^T u(\tau)^2 \, d\tau \quad . \tag{2.4.11}$$

i.e. $$\| W^* W \| \leq \int_0^T \left(\int_0^t |W(t,\tau)|^2 \, d\tau \right) dt \; . \tag{2.4.12}$$

While this is sometimes a crude bound, it illustrates the dependence of the norm on the interval T, and T should be no larger than necessary, four times the optimal response decay time being a reasonable upper limit for accurate results.

Example 2.4.2:

$$\min_u \; J = \sum_{j=0}^{9} y_k^2 + r \, u_k^2$$

for the system of Figure 2.4.2.

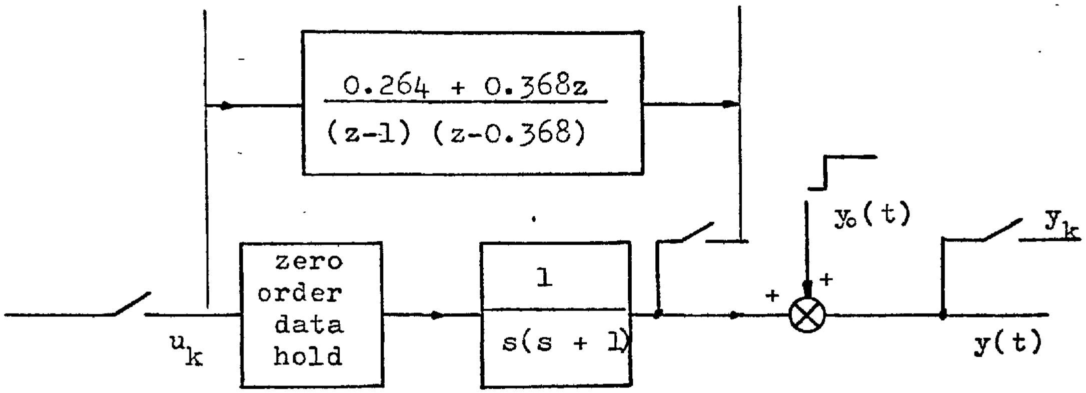


fig. 2.4.2

The computed trajectories are shown in Figure 2.4.3 for a range of r. The response of the continuous plant between sampling intervals is calculated by means of the modified z-transform. As r becomes smaller, the condition number $R + W^* Q W$ becomes larger, and more iterations are required. The optimal response for one value of r is used as the initial approximation for the next lower value. (This is true for all the examples in Section 2.4).

A unique solution for optimal u exists for all $r \neq 0$, but for $r = 0$, u_9, the final control value, may be anything at all, without affecting the cost. However, $u_9 \neq 0$ will be a poor solution, since

(a) it is desirable to make $u(r)$ continuous at the origin, and

(b) unless $u_9 \simeq 0$, the control trajectory on the finite interval will not approximate the infinite time optimal control.

The response for $r = 0$ (with $u_9 = 0$ for uniqueness) is to be expected from engineering intuition. Since $W_o = 0$ no amount of control can make $y_o = 0$. However, after this interval it is possible to make $y_k = 0$, for all $k > 0$.

i.e. $$y(z) = 1 \quad .$$

Hence $$u(z) = -\frac{(z-1)(z-0.368)}{0.264+0.368z} \cdot \frac{1}{z-1}$$

$$= -\frac{z-0.368}{0.264+0.368z}$$

which corresponds to the computed control. Consideration of the continuous output shows that $r = 0$ does not produce a good system, according to normal servo design criteria.

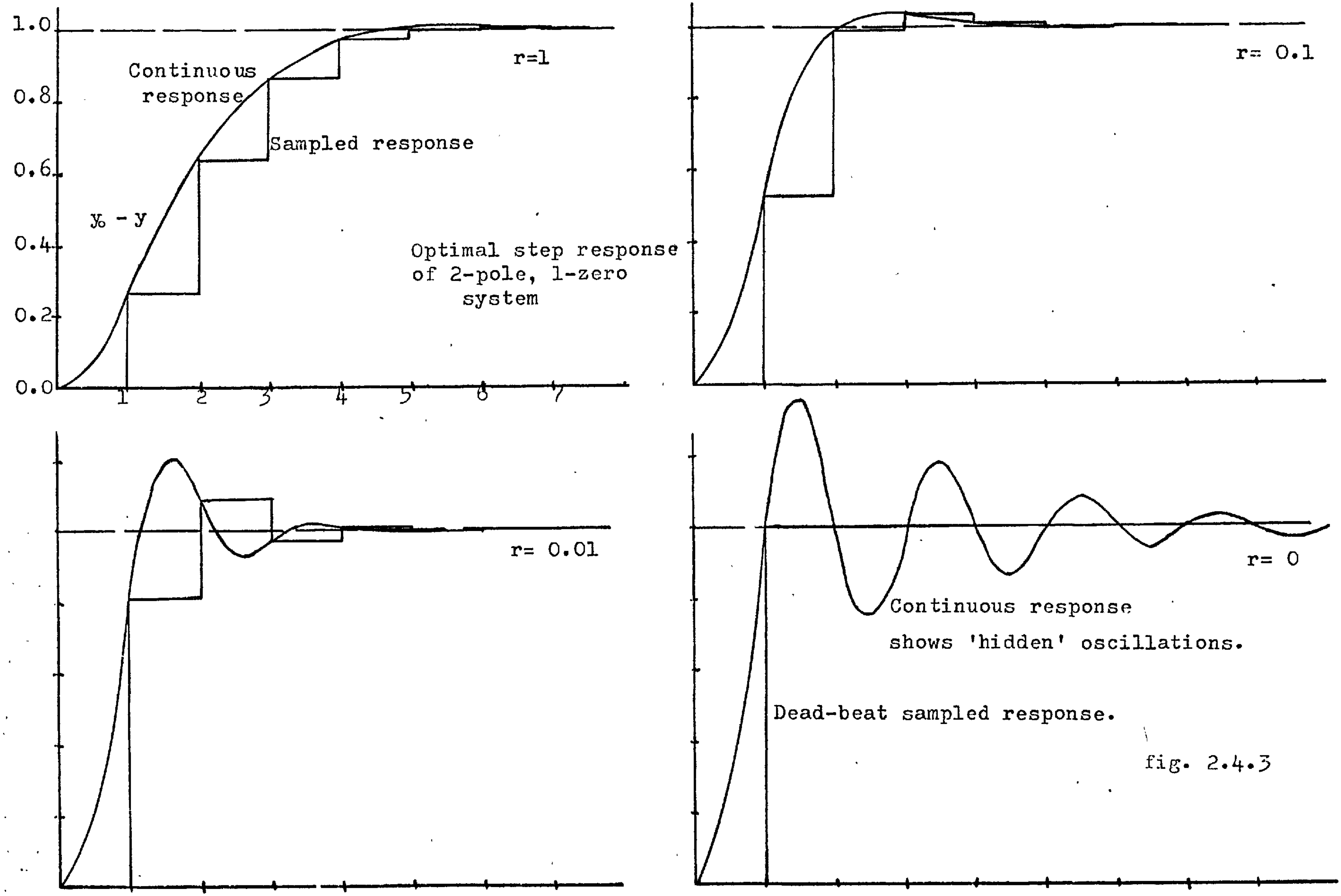
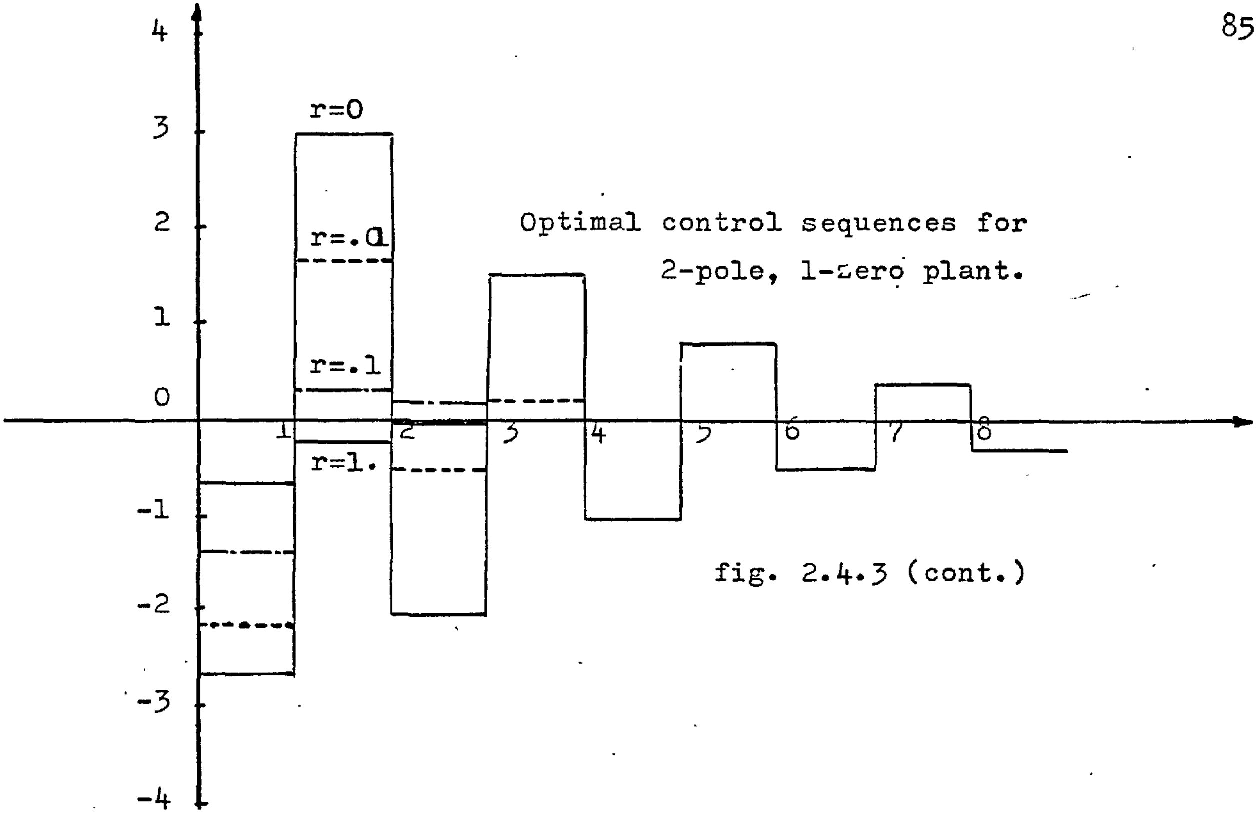

1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
$y_0 - y$
Continuous response
Sampled response
r=1
Optimal step response of 2-pole, 1-zero system
r= 0.1
r= 0.01
r= 0
Continuous response shows 'hidden' oscillations.
Dead-beat sampled response.

fig. 2.4.3



Optimal control sequences for 2-pole, 1-zero plant.

fig. 2.4.3 (cont.)

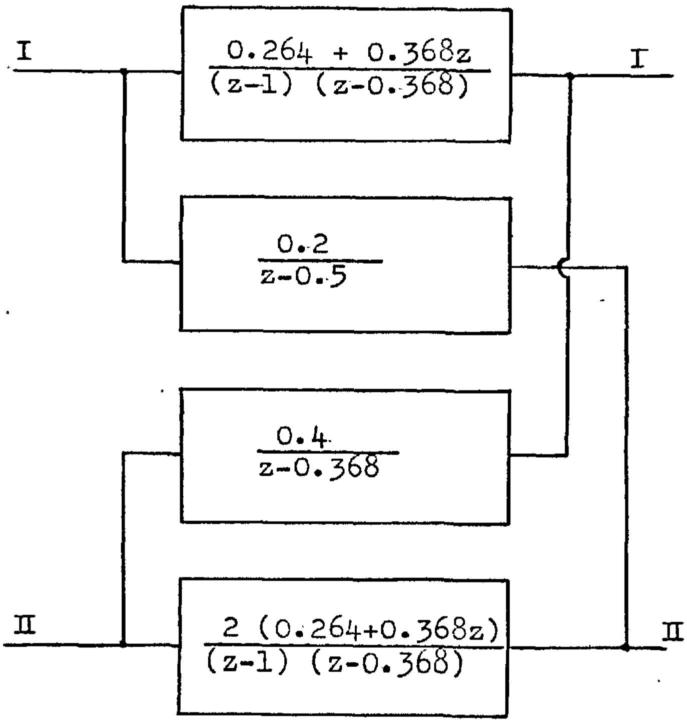

I
$\frac{0.264 + 0.368z}{(z-1)(z-0.368)}$
I
$\frac{0.2}{z-0.5}$
$\frac{0.4}{z-0.368}$
II
$\frac{2(0.264+0.368z)}{(z-1)(z-0.368)}$
II

fig. 2.4.4

Example 2.4.3: We require to minimise

$$J = \sum_{k=0}^{19} y_k^T y_k + u_k^T (r I) u_k$$

for the multivariable plant of Figure 2.4.4. Graphs similar to Figure 2.4.3 are plotted in Figure 2.4.5.

Example 2.4.4: The last three examples all possessed z-transforms which were simple rational functions, and hence amenable to the state-space treatment to be described in Section 2.8. However, we now consider the plant shown in Figure 2.4.6, which does not have a finite recurrence relation representation.

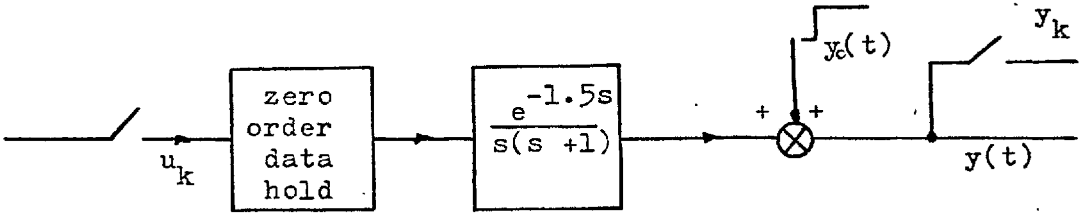
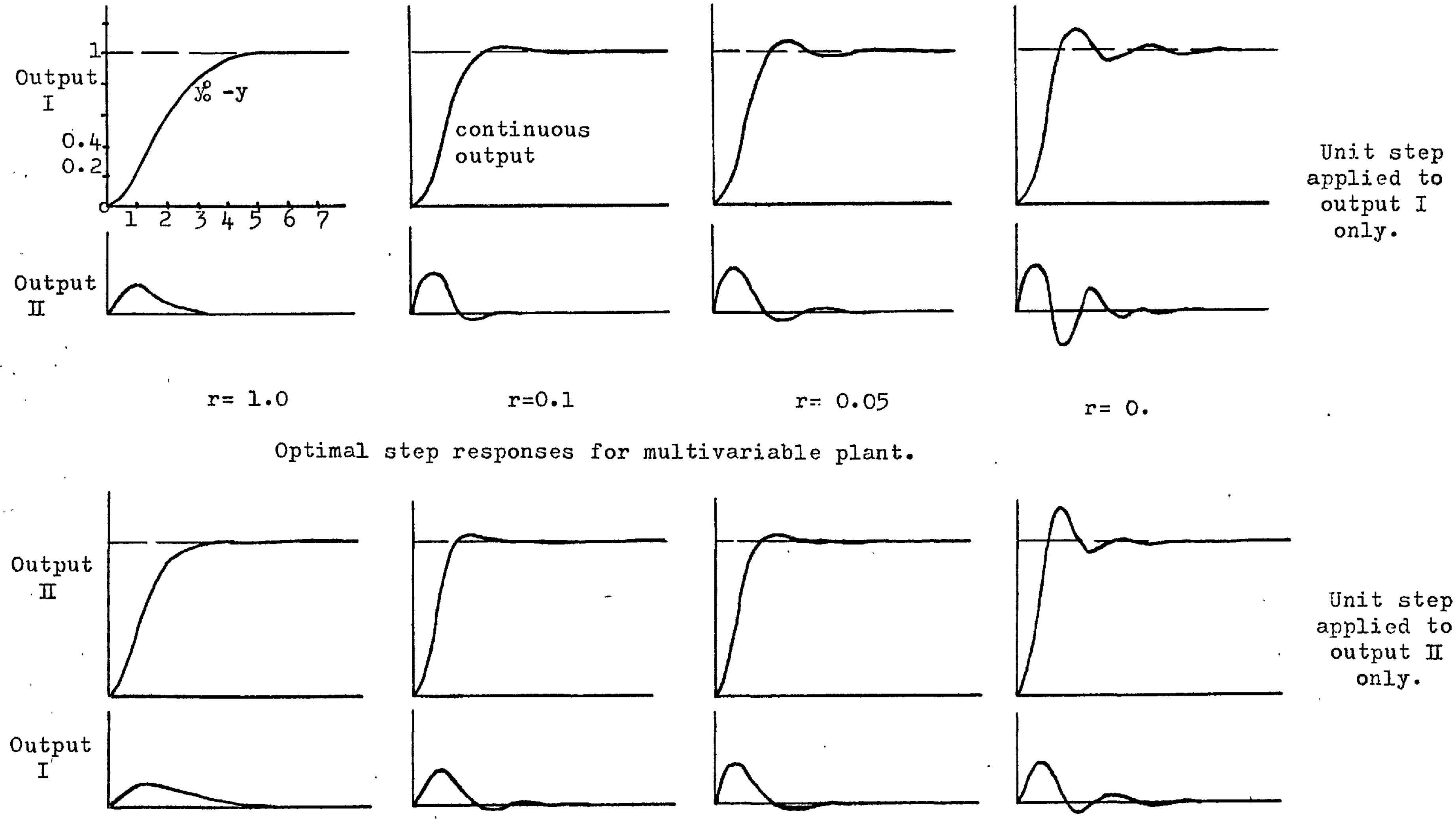


fig. 2.4.6

The dimension of the state space is irrelevant to our algorithms, and the computed optimal trajectories are plotted in Figure 2.4.7 for the same cost function and disturbance as example 2.4.2. The plant is non-minimum phase, since its inverse is anticipatory by more than one sample interval.

Example 2.4.5: Another non-minimum phase plant is the rational plant shown in Figure 2.4.8.



r= 1.0

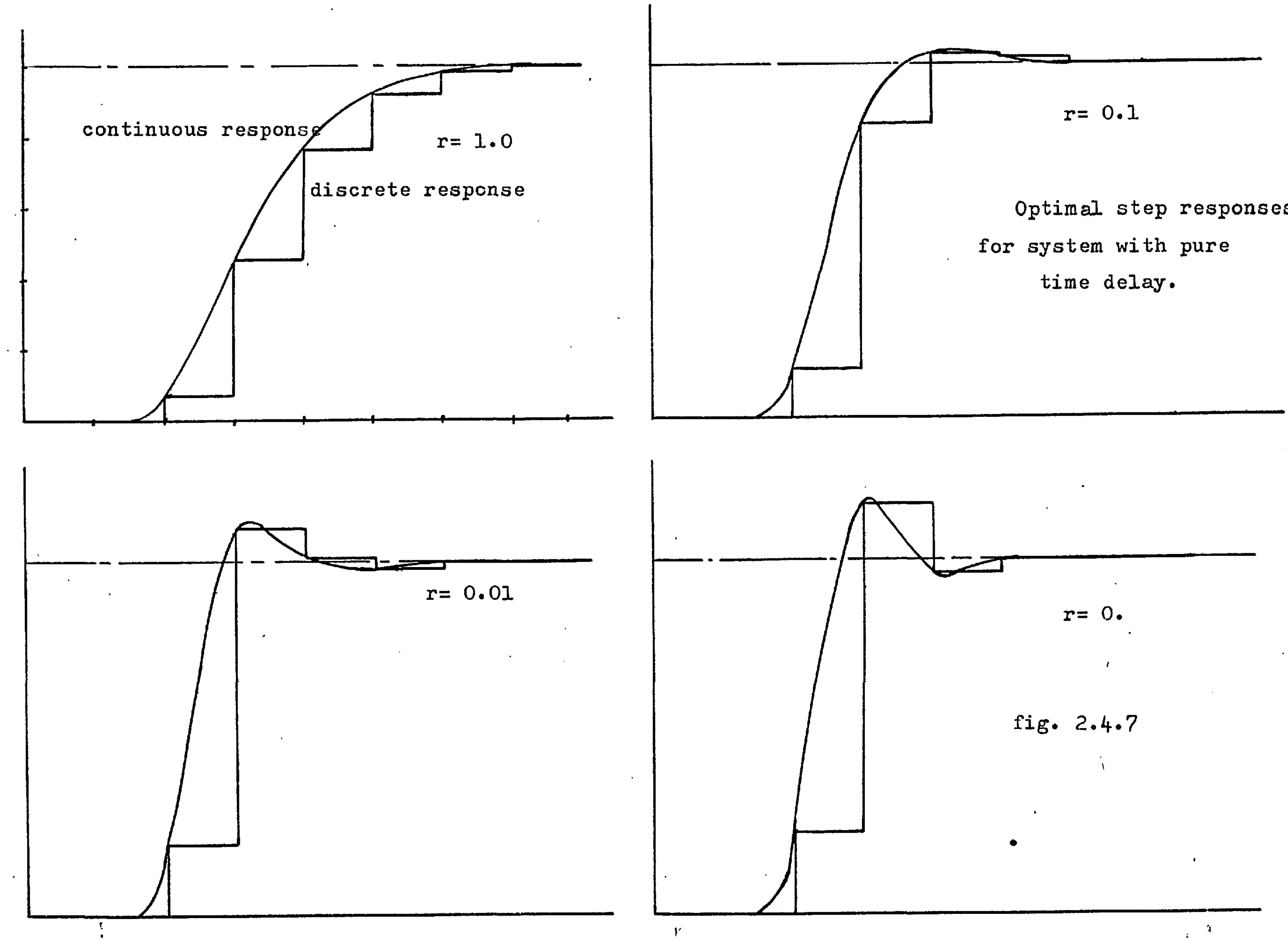
r=0.1

r= 0.05

r= 0.

Optimal step responses for multivariable plant.

fig. 2.4.5



Optimal step responses for system with pure time delay.

fig. 2.4.7

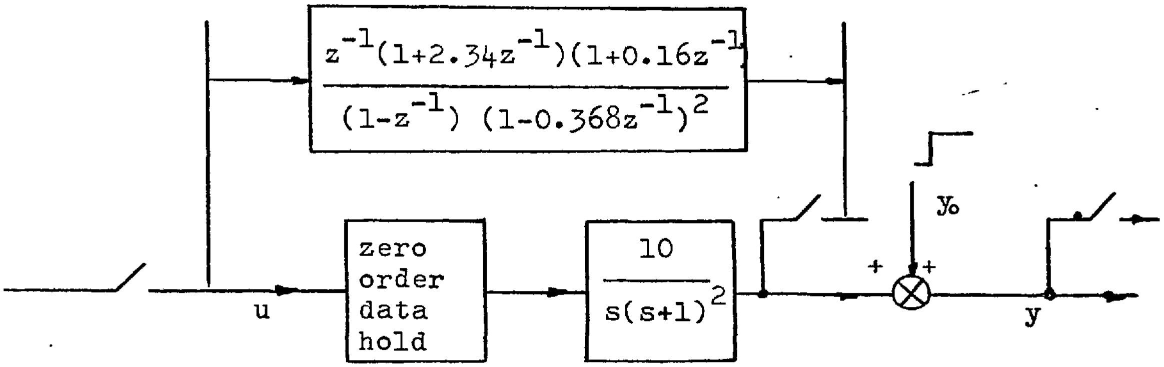
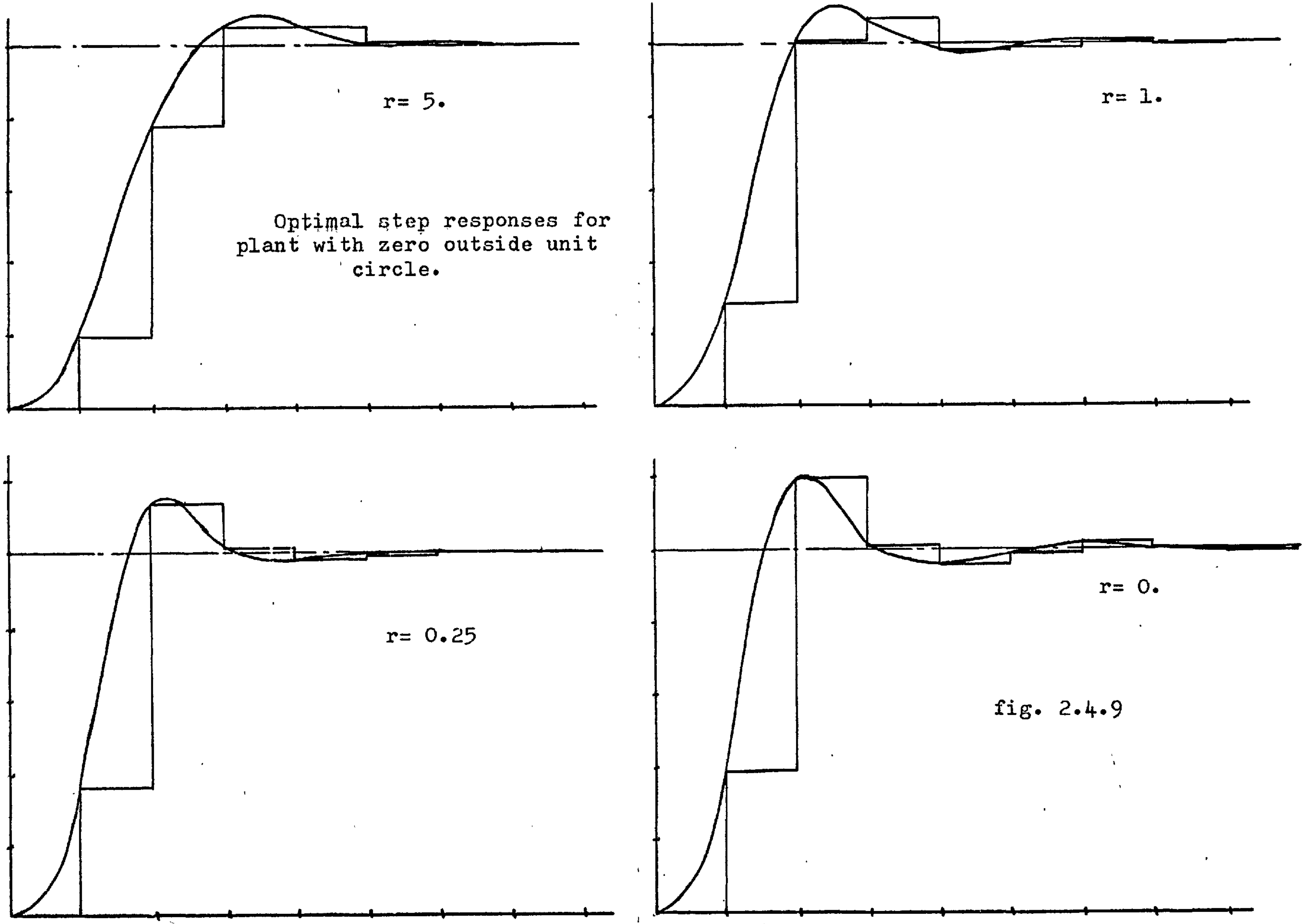

$\frac{z^{-1}(1+2.34z^{-1})(1+0.16z^{-1})}{(1-z^{-1})\ (1-0.368z^{-1})^2}$
zero order data hold
$\frac{10}{s(s+1)^2}$
u
y_0
+
+
y

fig. 2.4.8

The z-transform represents an N.M.P. plant, since the zero at $z = -2.34$ lies outside the unit circle, even though the original continuous plant is minimum phase. The N.M.P. property is introduced by the sampling operation. The results for the cost function of example 2.4.2 are plotted in Figure 2.4.9.

The interesting lesson from both the examples 2.4.4-5 is that, for $r = 0$, the type of dead-beat response obtained in example 2.4.2 did not result. Of course, for the pure-delay example 2.4.4, the system could not possibly respond until at least the second time instant, but after that the response may have been expected to be dead-beat. In fact, for example 2.4.5, the dead-beat response is indeed the true optimum and produces a lower cost than that calculated (1.0 vs ~1.4), and the conjugate gradient method has not converged to the correct answer. If a dead-beat response is assumed, and the control calculated as for example 2.4.2, then



Optimal step responses for plant with zero outside unit circle.

fig. 2.4.9

$$\hat{u}_1(z) = \frac{(1 - 0.368z^{-1})^2}{(1 + 2.34z^{-1})(1 + 0.16z^{-1})}$$

which is unstable due to the pole outside the unit circle. However, the time response restricted to any bounded interval is bounded, and hence gives the strict optimum control. However, it is not a useful type of control whereas the control calculated by the conjugate gradient algorithm is stable, and hence applicable in an engineering design. In fact, the control that we have computed corresponds to

$$\hat{u}_2(z) = \frac{(1 - 0.368z^{-1})^2}{(1 + 2.34z)(1 + 0.16z^{-1})}$$

i.e., the unstable pole has been replaced by its reciprocal with respect to the unit circle. This behaviour can be interpreted as an extreme instance of "ridge phenomenon". A plot of $\| \hat{u} \|$ vs r looks similar to the graph of Figure 2.4.10.

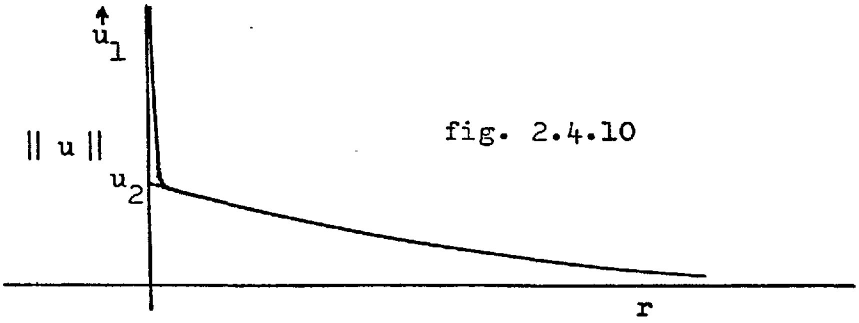


fig. 2.4.10

The knee of the curve becomes sharper and closer to the axis as the time interval is increased.

For a two-dimensional problem this state of affairs is simulated by highly eccentric contours, as in Figure 2.4.11.

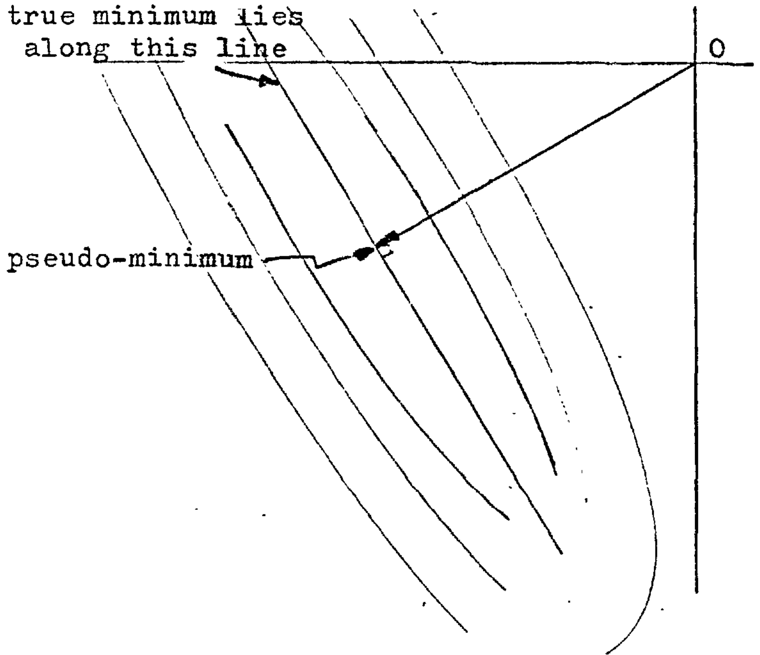


fig. 2.4.11

Due to numerical inaccuracies (which may be very small), the conjugacy property is lost and the search vectors are not quite in the exact directions. However, the cost for moving a small distance off the true conjugate direction may be several orders of magnitude larger than for large distances along the true direction, and hence only very small steps are taken around the pseudo-minimum. For

$$W^* y = 0$$

or

$$W^* W u = - W^* y_o$$

then

$$u = -(W^* W)^+ W^* y_o = - W^+ y_o$$

where + denotes pseudo-inverse. However, apart from the obvious singularity of W due to $W_o = 0$, the non-minimum phase effect approximately decreases the rank further; i.e. the cost contours (all Figure 2.4.11 for a two-dimensional representation) are almost degenerate (parallel straight lines in two-dimensions) in which case the control with minimum norm is found. From an engineering viewpoint, the interpretation of the small eigenvalues of $W^* W$ as zero gives a good stable design, and makes gradient algorithms preferable to direct inversion.

Example 2.4.7: A final example is the continuous system

$$y(t) = y_o(t) + \int_o^t W(t - \tau)\, u(\tau)\, d\tau$$

where $$y_o(s) = \frac{1}{s}$$

$$W(s) = \frac{e^{-\sqrt{s}}}{s}$$

i.e. $$W(t) = \operatorname{erfc}\left(\frac{1}{2\sqrt{t}}\right)$$

This could arise in an idealised model of the engineering system shown in Figure 2.4.12.

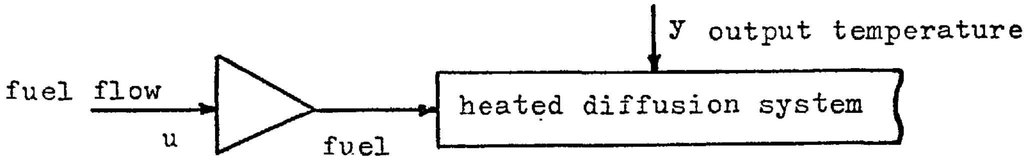


fig. 2.4.12

The weighting function is plotted in Figure 2.4.13, and the corresponding $W(j\omega)$ on the Nyquist diagram in Figure 2.4.14. Then with cost function

$$J = \int_0^{10} q\, y^2 + u^2\, dt$$

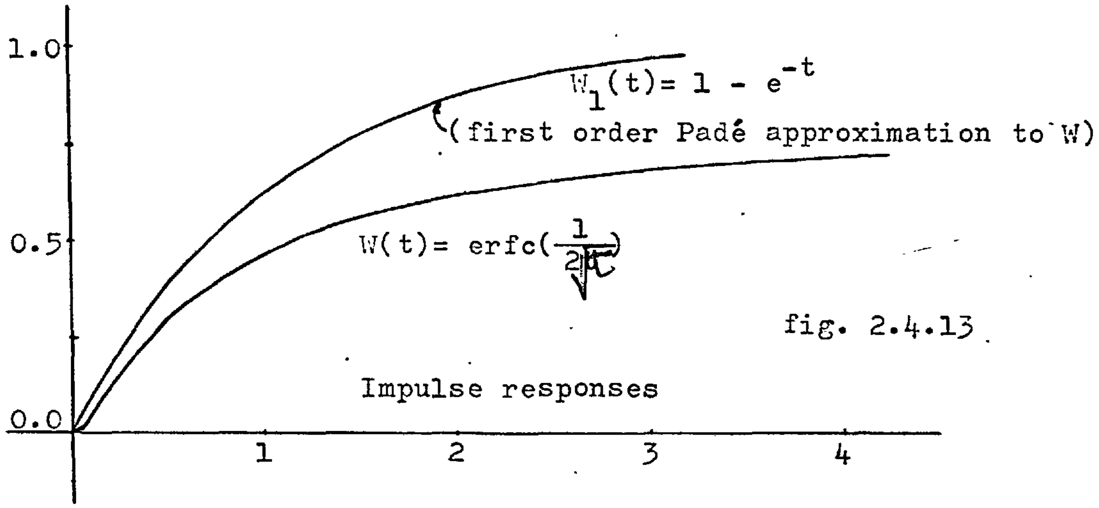
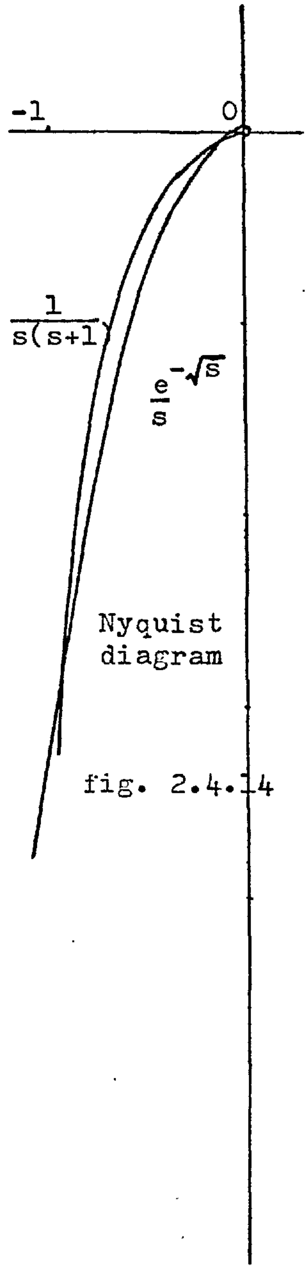


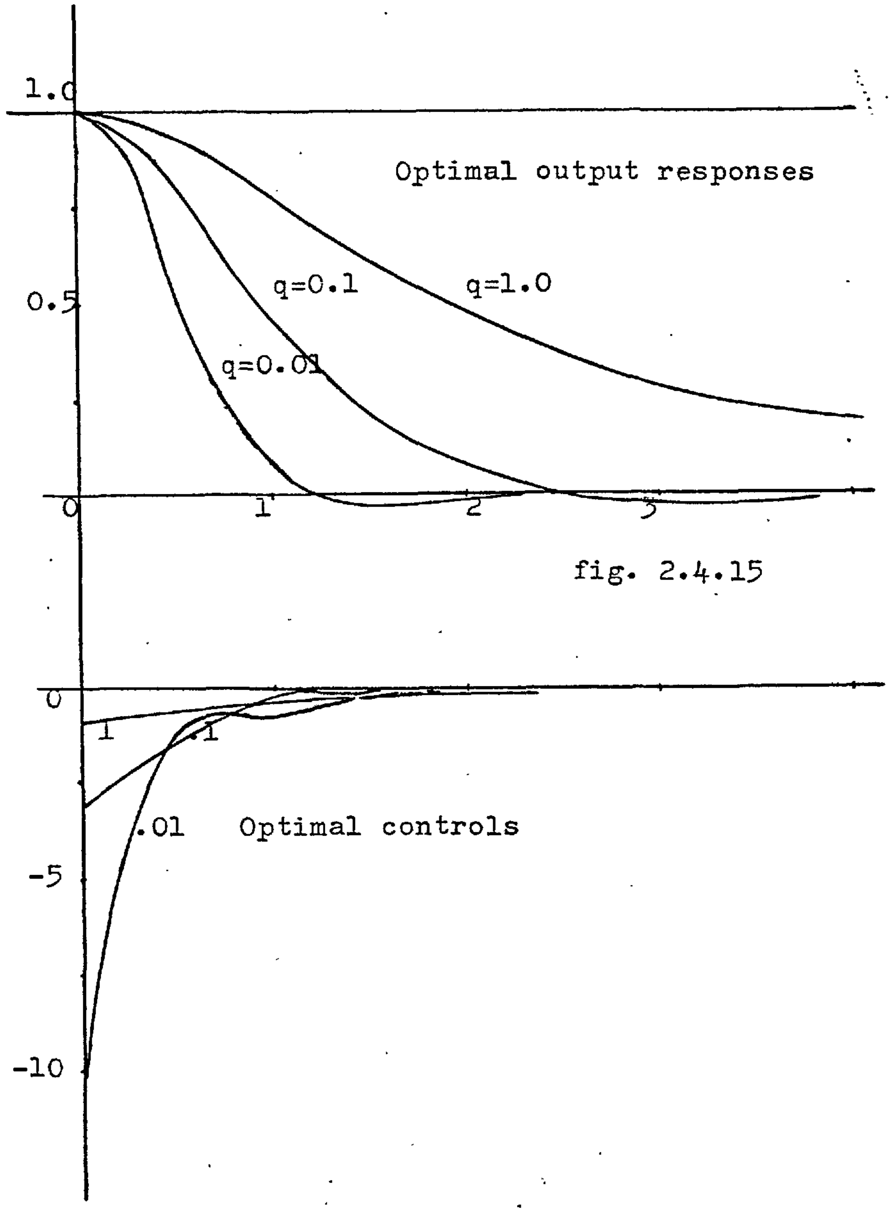
fig. 2.4.13

Impulse responses



Nyquist diagram

fig. 2.4.14



Optimal output responses

fig. 2.4.15

Optimal controls

the optimal trajectories have been computed and plotted in Figure 2.4.15. The integration procedure uses a fourth order Newton-Coates formula.

This method is considered excellent for calculating optimal trajectories for linear systems, especially where the weighting function is only known empirically. No state space approximations or identification is required. However, no structural configuration in feedback form is implied, and this has yet to be considered.

2.5 Direct Solutions and Return Difference

The algorithms developed in Section 2.3 are essentially iterative. Methods of solving the conditions of optimality explicitly are now investigated. One such method is the eigenvalue-eigenvector expansion mentioned in Section 2.2. The case of W being a causal operator is of particular interest. Causality is introduced by considering the Hilbert spaces $\mathcal{H}_u$ and $\mathcal{H}_y$ as subspaces of larger spaces H_u and H_y. We will change our notation and say $u \in \mathcal{H}_u^+$, $y \in \mathcal{H}_y^+$, where

$$\mathcal{H}_u^+ \oplus \mathcal{H}_y^- = H_u \qquad 2.5.1$$

and

$$\mathcal{H}_y^+ \oplus \mathcal{H}_y^- = H_y \quad . \qquad 2.5.2$$

For example, if $\mathcal{H}_u^+ = L_2 [0, \infty)$, then

$$H_u = L_2(-\infty, \infty) .$$

The operator W is causal, so, for all $u_+ \in \mathcal{H}_u^+$

$$Wu_+ \in \mathcal{H}_y^+$$

and $W \in \mathcal{J}_+(H_u, H_y)$. We shall assume that the operators $Q, R \in \mathcal{J}_+ \cap \mathcal{J}_-$ i.e. are instantaneous .

From Theorem 1.8.3, W^* is completely anticipatory from the space H_u into H_y. Since u can only be chosen on the space $\mathcal{H}_u^+$, the optimality condition 2.2.3 becomes

$$R\,u + \pi_+ (W^* Q\,Y) = 0 \qquad 2.5.3$$

We shall use an abbreviated notation for this projection operation

$$R\,u + [W^* Q\,y]_+ = 0 \qquad 2.5.4$$

or $$R\,u + W^* Q\,y - [W^* Q\,y]_- = 0 . \qquad 2.5.5$$

Put $$v = -[W^* Q\,y]_- . \qquad 2.5.6$$

Then $$v \in \mathcal{H}_u^-$$

and $$R\,u + W^* Q\,y + v = 0 .$$

But $$y = y_o + W\,u , \qquad 2.5.7$$

So $$(R + W^* Q\,W)\,u + W^* Q\,y_o + v = 0 . \qquad 2.5.8$$

A factorisation of the self-adjoint operator $R + W^* Q W$ is now assumed. Consider a transformation F from $\mathcal{H}_u^+$ into $\mathcal{H}_u^+$ that is causal and bounded, and whose inverse F^{-1} exists from $\mathcal{H}_u^+$ into $\mathcal{H}_u^+$ and is also bounded.

Theorem 2.5.1: Under the above conditions, F^{-1} is necessarily causal.

Proof: F is causal, so $v \in \mathcal{H}_u^+ \;\forall\; u \in \mathcal{H}_u^+$ where

$$v = F u$$

But F is invertable, so for any $v \in \mathcal{H}_u^+$, $\exists$ a $u \in \mathcal{H}_u^+$ such that

$$u = F^{-1} v$$

i.e. $F^{-1}$is causal.

Then we assume it is possible to find such an F so that

$$F^* F = R + W^* Q W \qquad 2.5.9$$

From Theorem 1.6.7, regardless of whether F^{-1} exists, all F which obey 2.5.9 differ by an isometric transformation. However

Theorem 2.5.2: If F satisfies 2.5.9, and has a bounded inverse, then all such F differ by a unitary transformation.

Proof: If $F^* F = R + W^* Q W$, then all F with this property differ by an isometric transformation (Theorem 1.6.7.).

i.e. $F = U F_1$,

where $U^* U = I$.

But F_1^{-1} exists by assumption, as does F^{-1} .

So $U = F F_1^{-1}$ which is bounded

and $U^{-1} = F_1 F^{-1}$ which is also bounded.

i.e. U is unitary.

With these assumptions, and a particular F, 2.5.8 becomes

$$F^* F u + W^* Q y_o + v = 0 \qquad 2.5.10$$

$$F u + F^{*-1} W^* Q y_o + F^{*-1} v = 0 \qquad 2.5.11$$

Now both sides of 2.5.11 are projected into $\mathcal{H}_u^+$. Since F is causal, $F u \in \mathcal{H}_u^+$. F^{*-1} is purely anticipatory, so

$$[F^{*-1} v]_+ = 0 \ . \qquad 2.5.12$$

Hence

$$F u + [F^{*-1} W^* Q y_o]_+ = 0 \qquad 2.5.13$$

$$u = - F^{-1} [F^{*-1} W^* Q y_o]_+ \ . \qquad 2.5.14$$

We digress to discuss a particular interpretation of these equations for time-invariant continuous systems. In this case, $W^* y$ is interpreted as the vector

$$z(t) = \int_t^\infty W^T(\tau - t)\, y(\tau)\, d\tau \qquad 2.5.15$$

and $z \in \mathcal{H}_u^+$ if $t \in [o, \infty)$. However, if $z(t)$ is defined $\forall$ $t \in (-\infty, \infty)$, then we are interested in calculating $[z]_+$. By Parseval's theorem

$$\int_o^\infty a^T(t)\, b(t)\, dt = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} a^T(-s)\, b(s)\, ds\,. \qquad 2.5.16$$

Hence $$z(t) = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} W^T(-s)\, e^{st}\,.\, y(s)\, ds\,. \qquad 2.5.17$$

By assumption $y(t) \in L_2^p[o, \infty)$ and hence $y(s)$ is analytic in the right half s plane, whereas $W^T(-s)$ is analytic in the left half s plane. Equation 2.5.17 resembles an inversion integral, but only the stable part is counted, since the implied contour integral excludes the right half plane singularities. Hence, if

$$W^T(-s)\, y(s) = a(s) + b(s)$$

$$= [W^T(-s)\, y(s)]_+ + [W^T(-s)\, y(s)]_-$$

where $a(s)$ is stable and $b(s)$ is purely unstable, then

$$z(s) = a(s) = [W^T(-s)\, y(s)]_+\,. \qquad 2.5.18$$

For bilateral Laplace transforms, there is an ambiguity of interpretation, which, however, does not affect our argument. For example, a transform

$$F(s) = \frac{s + 1}{1 - s} \quad .$$

may be interpreted as either unstable, but causal, or anticipatory, but stable. However, if we require causality and boundedness, then only left half plane singularities can be considered. Similarly, for $F^{-1}(s)$ to be causal and bounded, requires F to be minimum phase.

Example 2.5.1: Minimise

$$J = \int_0^\infty y^2 + u^2 \, dt$$

where
$$y(s) = \frac{1}{s + 1} + \frac{1}{s + 1} u(s) \quad . \qquad 2.5.19$$

$$R + W^* Q W = 1 + \frac{1}{- s + 1} \cdot \frac{1}{s + 1}$$

$$= \frac{2 - s^2}{(1 + s)(1 - s)}$$

$$= (\frac{\sqrt{2} - s}{1 - s})(\frac{\sqrt{2} + s}{1 + s}) \quad . \qquad 2.5.20$$

So
$$F = \frac{\sqrt{2} + s}{1 + s}$$
, which has a stable inverse.

Then
$$[F^{*-1} W^* Q y_o]_+ = [\frac{1 - s}{\sqrt{2} - s} \cdot \frac{1}{- s + 1} \cdot \frac{1}{s + 1}]_+$$

The projection operation simply takes these partial fractions whose poles are in the left half s plane.

i.e. $$[F^{*-1} W^{*} Q y_o]_+ = \frac{1}{\sqrt{2}+1} \cdot \frac{1}{s+1} \quad .$$

Then $$u(s) = \frac{-1+s}{\sqrt{2}+s} \cdot \frac{1}{\sqrt{2}+1} \cdot \frac{1}{s+1}$$

$$= \frac{-(\sqrt{2}-1)}{\sqrt{2}+s} \quad . \qquad 2.5.21$$

From 2.5.19, $$y(s) = \frac{1}{\sqrt{2}+s} \quad . \qquad 2.5.22$$

Transforming to the time domain

$$u(t) = -(\sqrt{2}-1)e^{-\sqrt{2}t} \quad ,$$

$$y(t) = e^{-\sqrt{2}t} \quad .$$

i.e. $$u(t) = -(\sqrt{2}-1)\, y(t) \; .$$

A similar interpretation in terms of the z-transform can be obtained for sampled systems, where the left half s plane maps into the interior of the unit circle in the z plane.

We now assume that a solution of the optimal control problem exists and investigate the spectral factorisation of $R + W^*QW$. We choose y_o to be initial conditions of the system.

i.e. $$y_o = [W\alpha]_+ \quad , \qquad 2.5.23$$

where $$\alpha \in \mathcal{H}_u^- \quad . \qquad 2.4.24$$

For systems on $L_2(l_2)$, this implies that the only initial conditions on the system are those that entered via the control input prior to time zero. If the system is controllable, then all initial conditions can be reached by an $\alpha(t)$, where $\alpha(t)$ is bounded and lies on a finite interval $[-t_1, 0]$.

From 2.5.8,

$$(R + W^* Q W)u + W^* Q y_o + v = 0 .$$

$$(R + W^* Q W)u + [W^* Q W \alpha]_+ + v = 0 ,$$

or $$(R + W^* Q W)u + W^* Q W \alpha + v_1 = 0$$

where $$v_1 \in \mathcal{H}_u^- .$$

i.e. $$(R + W^* Q W)u + (R + W^* Q W)\alpha + v_2 = 0 \qquad 2.5.25$$

where $$v_2 \in \mathcal{H}_u^- . \qquad 2.5.26$$

Now u depends linearly on α. We propose a solution

$$T u = -[T \alpha]_+ \qquad 2.5.27$$

where T is a causal operator to be determined and T^{-1} exists.

Then $$T u + T \alpha = \gamma \in \mathcal{H}_u^- \qquad 2.5.28$$

i.e. $$(R + W^* Q W)T^{-1}\gamma = v_2 \qquad 2.5.29$$

But γ, $v_2 \in \mathcal{H}_u^-$ but otherwise arbitrary, depending on α.

$$\left.\begin{aligned} \text{Hence} \quad (R + W^{*} Q\, W)T^{-1} &\in \mathcal{J}_- \\ &= S \;. \end{aligned}\right\} \qquad 2.5.30$$

Note that T^{-1} must be bounded, since $u \in \mathcal{H}_u^{+}$.

Hence $$R + W^{*} Q\, W = S\, T \;. \qquad 2.5.31$$

But $R + W^{*} Q\, W$ is self-adjoint .

i.e. $$T^{*} S^{*} = S\, T \;. \qquad 2.5.32$$

i.e. $$S^{*} T^{-1} = T^{*-1} S \;. \qquad 2.5.33$$

But $S^{*} T^{-1} \in \mathcal{J}_+$ whereas $T^{*-1} S \in \mathcal{J}_-$; i.e. both operators belong to $\mathcal{J}_+ \cap \mathcal{J}_-$.

i.e. $$S^{*} T^{-1} = T^{*-1} S = V \;. \qquad 2.5.34$$

$$S = T^{*} V \;. \qquad 2.5.35$$

So $$(R + W^{*} Q\, W) = T^{*} V\, T \;. \qquad 2.5.36$$

If V is factorisable, then this solves the spectral factorisation problem. However, there is no need to factor V in order to solve the optimal control problem by the equation 2.5.14.

Now $T \in \mathcal{J}_+$ is invertible. So split T into $D \in \mathcal{J}_+ \cap \mathcal{J}_-$, and $H \in \mathcal{J}_+ \setminus \mathcal{J}_+ \cap \mathcal{J}_-$.

i.e. $$T = D + H \qquad 2.5.37$$

where D is invertible.

Then
$$T = D(I + G) \ .$$

i.e.
$$(R + W^* Q W) = (I + G)^* D^* V D (I + G)$$

$$= (I + G)^* M (I + G) \ . \qquad 2.5.38$$

Also
$$u = T^{-1}[T \alpha]_+$$

$$= (I + G)^{-1} D^{-1} D[(I + G) \alpha]_+$$

$$= (I + G)^{-1}[G \alpha]_+ \qquad 2.5.39$$

Re-define
$$T = I + G \qquad 2.5.40$$

and call T the optimal return difference. The optimal structure becomes

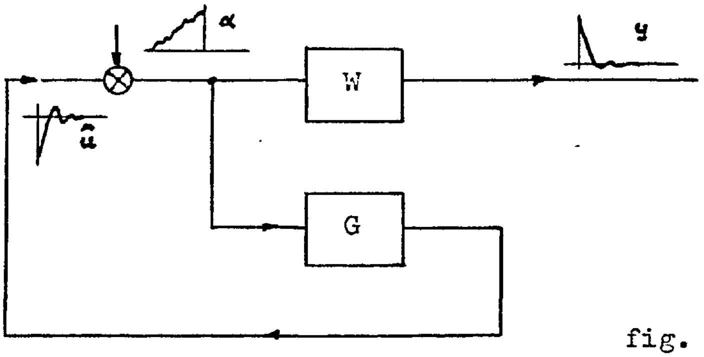


fig. 2.5.1

For continuous time systems, with infinite cost interval and time invariant operators, equation 2.5.38 can be interpreted as equations involving Laplace transforms and power spectra, in which case Kalman's result [K 2] is a corollary.

i.e. $$R + W^T(-s)\, Q\, W(s) \;=\; (I + G^T(-s))R\,(I + G(s)) \qquad 2.5.41$$

$M = R$ follows from the behaviour as s approaches infinity. Also, $G(s)$ must have the same singularities as $W(s)$.

Equation 2.5.38 can also be interpreted for discrete-time systems as z-transforms, in which case $M \neq R$. However, it is more general than transform equations, and can be interpreted as time-domain operations.

2.6 Contraction Algorithms

While a direct solution as discussed in section 2.5 is usually the best, for complicated problems factorisation becomes difficult to perform explicitly. We return to iterative methods of computing optimal controls. We have already discussed the iterative methods based on directions of descent, and these have proved very useful. In this section, algorithms based on contraction and the fixed point theorems discussed in Section 1.9 are discussed, and shown to have very useful properties, which sometimes give the engineer a better feel for the problem he is solving, than the gradient methods. They are also useful in that they can provide a solution of the spectral factorisation problem when applied to the Banach algebra of operators, rather than the Hilbert space of control trajectories. This will be discussed in Section 2.9. The methods presented in this section are not new to the theory of linear equations, but their interpretation and application to optimal control is believed to be new.

We wish to solve

$$(R + W^* Q W)u = f \tag{2.6.1}$$

Assume that the operator $R + W^* Q W$ can be decomposed into the sum

$$R + W^* Q W = A = D + A_+ + A_- \tag{2.6.2}$$

where

$$D \in \mathcal{J}_+ \cap \mathcal{J}_- \tag{2.6.3}$$

$$A_+ \in \mathcal{J}_+ \setminus (\mathcal{J}_+ \cap \mathcal{J}_-) \tag{2.6.4}$$

$$A_- \in \mathcal{J}_- \setminus (\mathcal{J}_+ \cap \mathcal{J}_-) \quad . \tag{2.6.5}$$

However, $R + W^* Q W$ is a self-adjoint operator on Hilbert space.

$$\therefore \quad \left.\begin{aligned} D^* &= D \quad , \quad \text{p.D.} \\ A_+ &= A_-^* \quad . \\ A_- &= A_+^* \quad . \end{aligned}\right\} \tag{2.6.6}$$

From equation 2.6.2,

$$(D + A_+ + A_-)u = f \quad . \tag{2.6.7}$$

Re-arrangement of this equation can lead to various well-known iterative algorithms for computing u.

Algorithm 2.6.1: Simple iteration; Jacobi's Method .

$$D\,u_{k+1} = f - (A_+ + A_-)\,u_k \quad . \qquad 2.6.8$$

Algorithm 2.6.2: One-step cyclic iteration; the method of Gauss-Seidel.

$$(D + A_+)\,u_{k+1} = f - A_-\,u_k$$

or

$$u_{k+1} = (D + A_+)^{-1}\ f - A_-\,u_k \quad . \qquad 2.6.9$$

Algorithm 2.6.3: Method of over-relaxation .

$$(D + \omega\,A_+)\,u_{k+1} = (1 - \omega)\,D\,u_k - \omega\,A_-\,u_k + \omega\,f$$

$$u_{k+1} = (D + \omega\,A_+)^{-1}\ (1 - \omega)\,D\,u_k - \omega\,A_-\,u_k + \omega\,f \ . \qquad 2.6.10$$

The parameter ω is called the relaxation factor and may depend on k in more complicated procedures. It is chosen to accelerate the convergence of the algorithm. Note that for $\omega = 1$, 2.6.10 reduces to 2.6.9.

These three algorithms have been well investigated in the literature, but mainly in the context of finite sets of linear equations. Collatz [C 1; p.223] describes these methods, and gives convergence criteria. In fact, all these methods have the form

$$u_{k+1} = T\, u_k + s \qquad 2.6.11$$

and converges if T is a contraction, by theorem 1.9.1.

i.e., if $$\|T\| < k < 1 , \qquad 2.6.12$$

the convergence rate being at least as fast as $(1 - k)^n$. A more specialised result which is useful for optimal control is a generalisation of theorem 3 of Collatz[C 1; p.228]. We extend this result to more general operators, and offer a different proof.

<u>Theorem 2.6.1</u>: If the operator $R + W^* Q W$ is self-adjoint, positive definite, and possesses a bounded inverse, then the Gauss-Seidel method, and, more generally, the relaxation method converge for $0 < \omega < 2$.

<u>Proof</u>:

$$u_{k+1} = (D + \omega A_+)^{-1}[(1 - \omega) D_{uk} - \omega A_- u_k + \omega f].$$

Hence $$T = (D + \omega A_+)^{-1}[(1 - \omega)D + \omega A_-] .$$

$$(D + \omega A_+)T = (1 - \omega) D - \omega A_- .$$

Multiply both sides by 2 .

$$(2 D + 2 \omega A_+)T = (2 - \omega) D - \omega D - 2 \omega A_- .$$

$$[(2 - \omega)D + \omega D + 2 \omega A_+]T = (2 - \omega) D - \omega D - 2 \omega A_- .$$

But $$A - A_- = D + A_+ .$$

$$[(2 - \omega)D + \omega(A - A_-) + \omega A_+]T = (2 - \omega)D - \omega(A - A_- - A_+) - 2\omega A_- .$$

$$[(2 - \omega)D + \omega A + \omega(A_+ - A_-)]T = (2 - \omega)D - \omega A + \omega(A_+ - A_-) . \qquad 2.6.13$$

Put $$S = (2 - \omega)D + \omega(A_+ - A_-) . \qquad 2.6.14$$

Then $$S + S^* = 2(2 - \omega)D + \omega(A_+ + A_- - A_- - A_+)$$

$$= 2(2 - \omega)D .$$

Since D is self-adjoint, P.D., so is

$$S + S^* \quad \text{for} \quad \omega < 2 .$$

Then $$(S + \omega A)T = (S - \omega A) \qquad 2.6.15$$

Put $$B = \omega A$$

Then B is P.D. self-adjoint, for $\omega > 0$.

i.e. for $0 < \omega < 2$, both $S + S^*$ and B are P.D., and

$$(S + B)T = S - B .$$

$$T = (S + B)^{-1}(S - B) . \qquad 2.6.16$$

Let λ be a (complex in general) eigenvalue of T, and x the associated eigenvector (which we assume exist).

$$(S + B)T x = (S + B) \lambda x$$

$$= (S - B)\,x \quad .$$

So $$\lambda S x + \lambda B x = S x - B x \quad .$$

$$(\lambda + 1) B x = (1 - \lambda) S x \quad .$$

$$(\lambda + 1) < x,\ B x > = (1 - \lambda) < x,\ S x > \qquad 2.6.17$$

$$= (1 - \lambda) < S^{*} x,\ x > \quad .$$

Take the complex conjugate of both sides .

$$(\bar{\lambda} + 1) < x,\ B x > = (1 - \bar{\lambda}) < x,\ S^{*} x > \quad . \qquad 2.6.18$$

Divide 2.6.17 by $1 - \lambda$, and 2.6.18 by $1 - \bar{\lambda}$.

$$\left[\left(\frac{\lambda + 1}{1 - \lambda}\right) + \left(\frac{\bar{\lambda} + 1}{1 - \bar{\lambda}}\right) \right] < x,\ B x > = < x,\ (S + S^{*}) x > \quad . \qquad 2.6.19$$

i.e. $$\frac{\lambda + 1}{1 - \lambda} + \frac{\bar{\lambda} + 1}{1 - \bar{\lambda}} > 0 \quad .$$

$$\frac{(\lambda + 1)(1 - \bar{\lambda}) + (\bar{\lambda} + 1)(1 - \lambda)}{|1 - \lambda|^{2}} > 0 \quad .$$

or $$\frac{2(1 - \lambda \bar{\lambda})}{|1 - \lambda|^{2}} > 0 \quad .$$

i.e. $$|\lambda| < 1 \quad \text{for} \quad 0 < \omega < 2 \ .$$

Hence $\|T\| < k < 1$, so the mapping is a contraction and convergence follows by theorem 1.9.1.

Collatz [C1] and Varga [V2] give many interesting properties of these algorithms including convergence and error bounds, and methods of choosing good values of ω. We present an example to show how these equations can be interpreted for control systems.

Example 2.6.1: We choose the same system as in Example 2.5.1.

$$J = \int_0^\infty y^2 + u^2 \, dt \ .$$

$$y(s) = \frac{1}{s+1}(1 + u(s)) \ .$$

Now $$(R + W^* Q W)(s) = 1 + \frac{\frac{1}{2}}{s+1} + \frac{\frac{1}{2}}{-s+1}$$

which corresponds to the operator splitting of 2.6.2, where

$$A_+ = \frac{\frac{1}{2}}{s+1} \ .$$

Also $$-W^* Q y_o = -\left[\frac{1}{-s+1} \cdot \frac{1}{s+1}\right]_+$$

$$= \frac{-\frac{1}{2}}{s+1} \ .$$

The relaxation algorithm becomes

$$\left(1 + \frac{\frac{\omega}{2}}{s+1}\right)u_{k+1} = (1-\omega)u_k - \omega\left[\frac{\frac{1}{2}}{1-s} u_k\right]_+ - \frac{\frac{1}{2}\omega}{s+1} \ .$$

But $$\left(1 + \frac{\frac{\omega}{2}}{s+1}\right)^{-1} = \left(\frac{s+1+\frac{\omega}{2}}{s+1}\right)^{-1}$$

$$= \left(\frac{2s+2+\omega}{2(s+1)}\right)^{-1}$$

$$= \frac{2(s+1)}{2s+2+\omega} \; .$$

Then the relaxation algorithm becomes

$$u_{k+1} = \frac{2(s+1)}{2s+2+\omega} \left\{ (1-\omega)u_k - \omega\left[\frac{\frac{1}{2}}{1-s} u_k \right]_+ - \frac{\frac{1}{2}\omega}{s+1} \right\} .$$

For any ω, the solution

$$u = -\left(\frac{\sqrt{2}-1}{s+\sqrt{2}}\right)$$

is a fixed point of the algorithm, as is easily checked.

(a) $\omega = 1$;

$$u_{k+1} = \frac{s+1}{2s+3} \left\{ \frac{1}{1-s} u_k \quad + \quad - \frac{1}{s+1} \right\}$$

$$u_o = -\frac{1}{s+1} \qquad\qquad u_o(t)\Big|_{t=0} = -1$$

$$u_1 = -\frac{\frac{3}{2}}{2s+3} \qquad\qquad u_1(t)\Big|_{t=0} = -0.75$$

$$u_2 = -\frac{135+18}{5(2s+3)^2} \qquad\qquad u_2(t)\Big|_{t=0} = -0.65$$

$$u_3 = -\frac{600s^2+1126s+954}{125(2s+3)^3} \qquad\qquad u_3(t)\Big|_{t=0} = -0.5$$

The time domain curves are plotted in Figure 2.6.1.

Restart algorithm:

(b) $\omega = 0.8;$

$$u_{k+1} = \frac{s+1}{s+1.4}\left\{ 0.2\, u_k - \left[\frac{0.4}{1-s} u_k\right]_+ - \frac{0.4}{s+1}\right\}$$

$$u_o = \frac{-\frac{1}{2}}{s+1} \qquad u_o(0) = -0.5$$

$$u_1 = \frac{-0.4}{s+1.4} \qquad u_1(0) = -0.4$$

$$u_2 = \frac{-(0.4133s + .5733)}{(s+1.4)^2} \qquad u_2(0) = -0.4133$$

The time domain trajectories are plotted in Figure 2.6.2.

It is seen that the method expands the closed-loop response into an infinite product of poles and zeros. The factor ω seems to give good convergence if it is chosen so that an approximation to the dominant closed loop mode appearsinthe denominator of the inverted operator.

A computer algorithm working in the time domain is very easy to implement. For example, for $L_2[o,\infty)$, the equation 2.6.9 becomes

$$R\, u(t) + \int_o^t S_1(t,\tau)\, u(\tau)\, d\tau + \int_t^\infty S_2(\tau, t)\, u(\tau)\, d\tau = f(t)$$

where $$S_1(t,\tau) = \int_t^\infty W^T(s, t)\, Q\, W(s,\tau)\, ds$$

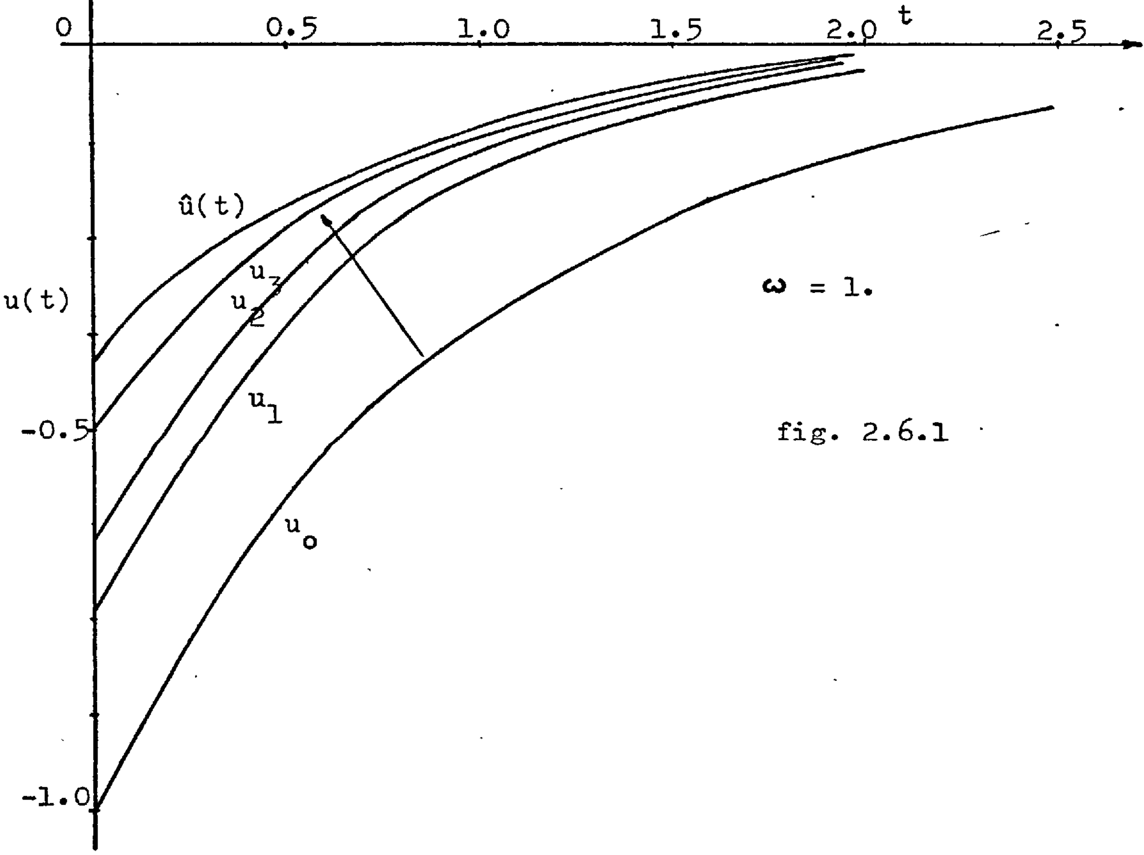


fig. 2.6.1

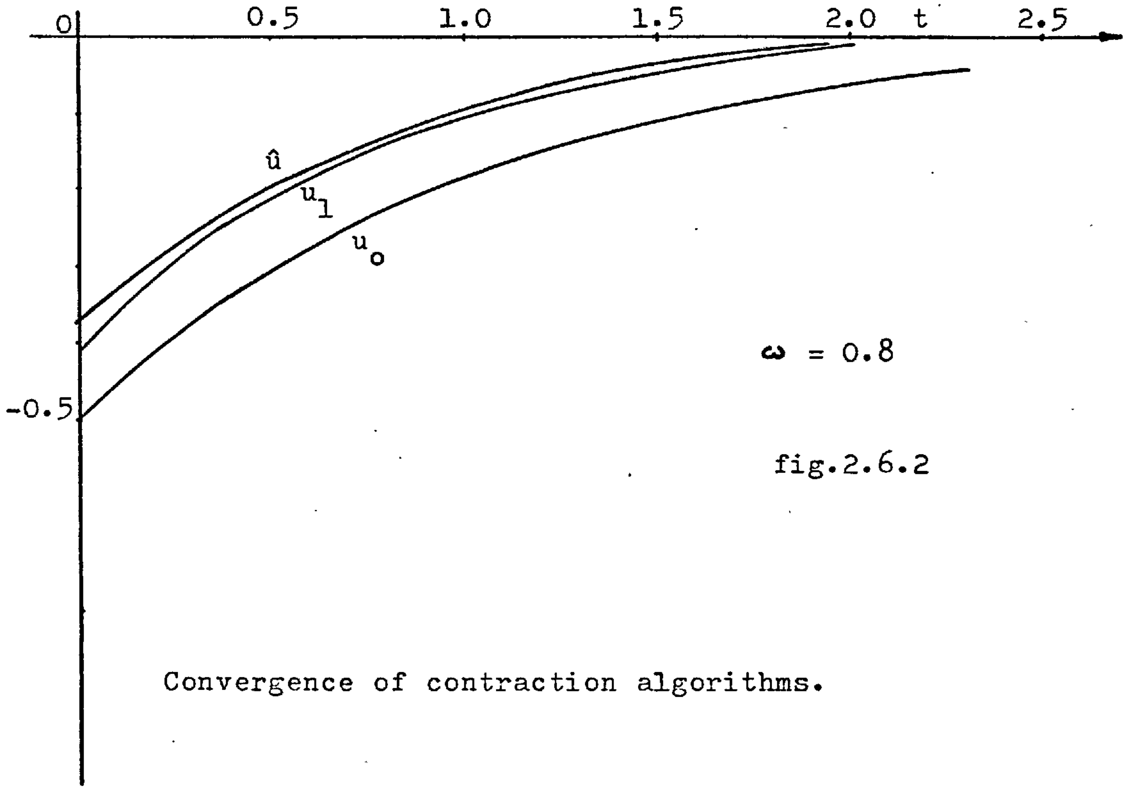


fig.2.6.2

Convergence of contraction algorithms.

and $$S_2(\tau,\, t) \;=\; S_1^{\,T}(t, \tau)\,.$$

First of all, the resolvent kernel of the equation

$$R\,u(t) \;+\; \omega \int_0^t S_1(t, \tau)\,u(\tau)\,d\tau \;=\; g(t) \qquad 2.6.21$$

is found, by iteration, and this scheme always converges for Volterra equations.

$$\text{i.e.} \quad u(t) \;=\; R^{-1}\,g(t) \;+\; \int_0^t K(t,\, \tau)\,g(\tau)\,d\tau\,. \qquad 2.6.22$$

$$\text{Then} \quad g(t) \;=\; f(t) \;-\; (1-\omega)\int_0^t S_1(t,\, \tau)u(\tau)d\tau - \int_1^\infty S_2(\tau,\, t)\,u(\tau)\,d\tau. \qquad 2.6.23$$

These iterative methods for solving integral equations, and general functional equations, have been neglected in the field of optimal control, but seem attractive computationally, since only integration is required. The main disadvantage of the method is that a feedback type structure is not predicted.

A basic operation to be performed is the inversion of the operator. $D \;+\; \omega\, A_+$. Does this inverse exist, and is it stable? This is answered by the following theorem, in the case of operators represented as Laplace transforms.

Theorem 2.6.2: If R, Q are P.D., P.S.D. symmetric matrices, such that

$$R + G(s) + G^T(-s) = R + W^T(-s)\, Q\, W(s) \qquad 2.6.24$$

then $(R + \omega G(s))^{-1}$ is stable for any allowable R and any $\omega > 0$.

Proof: Consider first the special case when $R = r$, a positive scalar, and $g(s)$ is a scalar transfer function.

Then $$2 \operatorname{Re} g(j\omega) = g(j\omega) + g(-j\omega) \quad . \qquad 2.6.25$$

i.e. $$2 \operatorname{Re} g(j\omega) = q\,|w(j\omega)|^2$$

$$> 0 \quad .$$

But $g(s)$ is stable, so $g(j\omega)$ lies entirely within the right half Nyquist plane, and so is stable for any gain whatever.

i.e. $$r^{-1}(1 + \frac{\omega}{r} g(s))^{-1} \quad \text{is stable}$$

or $$(r + \omega\, g(s))^{-1} \quad \text{is stable} \quad \forall\; r, \omega > 0 \,.$$

Now consider $(R + \omega G)^{-1}$.

$$G + G^* = W^* Q W \quad .$$

i.e. $$G + G^* \quad \text{is P.S.D.}$$

But if $\lambda(j\omega)$ is an eigenvalue of $G(j\omega)$, then $\lambda + \bar{\lambda}$ is an eigenvalue of $G + G^*$, where $\bar{\ }$ denotes complex conjugate. Therefore

$$\operatorname{Re}(\lambda) > 0 \quad .$$

But $$(R + \omega G) = R^{\frac{1}{2}}(I + \omega R^{-\frac{1}{2}} G R^{-\frac{1}{2}}) R^{\frac{1}{2}} .$$

where $R^{\frac{1}{2}}$ exists, P.D., since R is P.D.

Put $$H = R^{-\frac{1}{2}} G R^{-\frac{1}{2}} .$$

Then if $\mu(j\omega)$ is an eigenvalue of $H(j\omega)$, by the same reasoning as above

$$\mathrm{Re}(\mu) > 0 .$$

Now diagonalise H. i.e. we find an M, such that

$$H = M(\mathrm{diag}\ \mu_i)\ M^{-1} .$$

Then $$(R + \omega\ G) = R^{\frac{1}{2}}\ M\ (I + \omega(\mathrm{diag}\ \mu_i))\ M^{-1}\ R^{\frac{1}{2}} . \qquad 2.6.26$$

$(R + \omega\ G)^{-1}$ is stable if $\det(R + \omega\ G)$ is stable ;

i.e. if $\det(I + \omega\ \mathrm{diag}\ \mu_i)$ is stable.

But this will be stable if each loop

$$1 + \omega\ \mu_i\ (j\omega) \text{ is stable.}$$

However, $\mathrm{Re}\ \mu_i\ (j\omega) > 0$, so $\mu_i\ (j\omega)$ can never encircle the -1 point on the Nyquist diagram; in fact, it is stable for all $\omega > 0$.

Theorem 2.6.3: The iteration scheme

$$u_{k+1} = T\ u_k + s$$

derived from the equation

$$(R + W^{*} Q W) u + W^{*} Q y_o = 0$$

by algorithms 2.6.2-3, produces a decrease in cost at each iteration if T is a contraction with respect to the norm $\| x \| = < x, A x >^{\frac{1}{2}}$, where $A = R + W^{*} Q W$.

<u>Proof</u>: From equation 2.2.5, the change in cost in one iteration is

$$J_k = < 2 \varepsilon_k + A \Delta u_k, \Delta u_k > . \tag{2.6.27}$$

But

$$\varepsilon_k = (R + W^{*} Q W) u_k + W^{*} Q y_o \tag{2.6.28}$$

$$= M(u_k - T u_k - s)$$

$$= - M \Delta u_k$$

where

$$A = M(I - T) \tag{2.6.29}$$

and

$$M^{-1} = (I - T) A^{-1} .$$

So

$$\Delta u_k = - M^{-1} \varepsilon_k$$

$$= (T - I) A^{-1} \varepsilon_k . \tag{2.6.30}$$

$$\Delta J_k = < 2 \varepsilon_k + A T A^{-1} \varepsilon_k - \varepsilon_k, (T - I) A^{-1} \varepsilon_k >$$

$$= < \varepsilon_k + A T A^{-1} \varepsilon_k, A^{-1}(A T A^{-1} - I) \varepsilon_k >$$

$$= -\langle \varepsilon_k, A^{-1} \varepsilon_k \rangle - \langle A T A^{-1} \varepsilon_k, A^{-1} \varepsilon_k \rangle$$

$$+ \langle \varepsilon_k, A^{-1} A T A^{-1} \varepsilon_k \rangle + \langle A T A^{-1} \varepsilon_k, A^{-1} A T A^{-1} \varepsilon_k \rangle$$

$$= -\langle \varepsilon_k, A^{-1} \varepsilon_k \rangle + \langle A T A^{-1} \varepsilon_k, T A^{-1} \varepsilon_k \rangle .$$

Put $h = A^{-1} \varepsilon_k$.

$$\Delta J = -\langle A h, h \rangle + \langle A T h, T h \rangle . \qquad 2.6.31$$

However, by assumption,

$$\langle T h, A T h \rangle < \langle h, A h \rangle . \qquad 2.6.32$$

So there will be a reduction in cost at each iteration.

Corollary to Theorem 2.6.1: Under the conditions of Theorem 2.6.1, T is a contraction with respect to the norm $\| x \| = \langle x, A x \rangle^{\frac{1}{2}}$.

Proof: Using the notation of theorem 2.6.1

$$(S + B)T = S - B \qquad 2.6.33$$

where for $0 < \omega < 2$, $S + S^*$ and B are P.D.

Now $$ST + BT = S - B .$$

$$B(T + I) = S(I - T) .$$

Then $$(I - T^*)B(I + T) = (I - T^*)S(I - T) .$$

$$B - T^*B + BT - T^*BT = (I - T^*)S(I - T) .$$

Now operate on any non-zero vector h, and take the inner product with respect to h.

$$\langle h, Bh\rangle - \langle h, T^*Bh \rangle + \langle h, BTh\rangle - \langle h, T^*BTh \rangle$$

$$= \langle h, (I - T^*)S(I - T)h \rangle$$

$$\langle h, Bh \rangle - \langle h, T^*BTh \rangle = \langle h, (I - T^*)S(I - T)h \rangle$$

$$> 0 \qquad \text{for } 0 < \omega < 2 \ .$$

i.e. $$\langle h, Ah \rangle < \langle Th, ATh\rangle \ .$$

2.7 Unbounded Operators

It is often the case that the system operators are unbounded (unstable) and the unforced response y_o is also unbounded, and yet it is possible to find a $u \in \mathcal{H}_u$, such that

$$y = y_o + W u \in \mathcal{H}_y \ . \qquad 2.7.1$$

When this is so, u is said to stabilise the system. If such a u exists, then there will exist a $\hat{u} \in \mathcal{H}_u$, and a $\hat{y} \in \mathcal{H}_y$, which minimise the performance criterion

$$J = \langle y, Qy \rangle + \langle u, R u \rangle ,$$

providing W is closed and the domain of W is dense in $\mathcal{H}_u$. That is, we assume there exist vectors Δu, $\Delta y \in \mathcal{H}_u$, $\mathcal{H}_y$ respectively, such that

$$\Delta y = W \Delta u \qquad 2.7.2$$

and $\Delta u \in \mathcal{D}(W)$, the domain of W, which is a linear manifold in $\mathcal{H}^{\bullet}_u$, whose closure is all of $\mathcal{H}_u$. [RN 1; p.297]. The existence of $\hat{u}$ can be proved in a similar way to the bounded case. If there is no other u, which stabilise the system, than that hypothesised, then this is the unique optimal control. If there is more than one u, then suppose there exist u_1, $u_2 \in \mathcal{H}_u$, and y_1, $y_2 \in \mathcal{H}_y$, which satisfy 2.7.1. Then $\Delta u \in \mathcal{H}_u$ and $\Delta y \in \mathcal{H}_y$, and

$$\Delta y = W \Delta u \quad .$$

But W is linear, so its domain is a linear manifold. If Δu is restricted to $\mathcal{D}(W)$, the cost remains finite. As in the bounded case, we find a necessary and sufficient condition for optimality, and then show that $\hat{u}$ exists which satisfies this.

From Section 2.2

$$J = 2<\Delta u, R \hat{u} > + 2<\Delta y, Q \hat{y}> + <\Delta u, R \Delta u> + <\Delta y, Q \Delta y>$$

$$= 2<\Delta u, R \hat{u} > + <W\Delta u, Q \hat{y}> + <\Delta u, R \Delta u> + <W\Delta u, Q W\Delta u>.$$

Since $\mathcal{D}(W)$ is dense in $\mathcal{H}_u$, W^* can be defined [RN 1; p.299] and we obtain

$$J = \langle \Delta u, 2\hat{g} + (R + W^* Q W)\Delta u \rangle \quad . \qquad 2.7.3$$

Then we use

<u>Theorem 2.7.1</u>: If the linear transformation T is closed and its domain is dense in $\mathcal{H}$, the transformations

$$B = (I + T^*T)^{-1} , \qquad C = T(I + T^*T)^{-1} \qquad 2.7.4$$

are defined everywhere and bounded,

$$\|B\| \leqq 1 \qquad , \quad \|C\| \leqq 1 \qquad 2.7.5$$

and B is symmetric and positive.

<u>Proof</u>: Riesz and Nagy [RN 1; p.307].

Now consider the operator $S W R^{-\frac{1}{2}}$, where S is any operator such that

$$S^* S = Q \quad . \qquad 2.7.6$$

Q, and hence S, is assumed bounded, so $S W R^{-\frac{1}{2}}$ is closed, with domain dense in $\mathcal{H}_u$. (This implies some kind of controllability and observability which needs further research.)

Then $$B = (I + R^{-\frac{1}{2}} W^* Q W R^{-\frac{1}{2}})^{-1} \qquad 2.7.7$$

and $$C = S W R^{-\frac{1}{2}} (I + R^{-\frac{1}{2}} W^* Q W R^{-\frac{1}{2}})^{-1} \qquad 2.7.8$$

exist, are bounded, and have norms less than unity.

Hence $$R^{-\frac{1}{2}} B R^{-\frac{1}{2}} = \left[R^{\frac{1}{2}} \quad (I + R^{-\frac{1}{2}} W^{*} Q W R^{-\frac{1}{2}}) R^{\frac{1}{2}} \right]^{-1}$$

$$= (R + W^{*} Q W)^{-1} \qquad 2.7.9$$

is bounded, and

$$\| (R + W^{*} Q W)^{-1} \| \leqq \| R^{-1} \| \quad . \qquad 2.7.10$$

Similarly, $W(R + W^{*} Q W)^{-1}$ is bounded, and so is its adjoint

$$(R + W^{*} Q W)^{-1} W^{*} .$$

Now if $\hat{g} \neq 0$

then a change Δu from $\hat{u}$ is proposed, given by

$$\Delta u = - (R + W^{*} Q W)^{-1} \hat{g}$$

$$= - (R + W^{*} Q W)^{-1} R \hat{u} + (R + W^{*} Q W)^{-1} W^{*} Q \hat{y}$$

which is bounded, by virtue of the fact that all operators on the right are bounded, and $\hat{u} \in \mathcal{H}_u$, $\hat{y} \in \mathcal{H}_y$.

But $$\cdot J = < - (R + W^{*} Q W)^{-1} \hat{g} , \; 2\hat{g} - \hat{g} >$$

$$= - < (R + W^{*} Q W)^{-1} \hat{g} , \; \hat{g} >$$

$$\leqq 0$$

by Theorem 2.6.1 and equation 2.7.9. Therefore, unless $\hat{g} = 0$, $\hat{u}$ cannot be an optimum control. If such a $\hat{u}$ exists, then $\hat{g} = 0$

is also sufficient for a minimum, since $(R + W^* Q W)^{-1}$ is positive definite. The qu^stion of existence can now be investigated.

Consider the control

$$\hat{u} = -(R + W^* Q W)^{-1} W^* Q y_o . \qquad 2.7.11$$

By assumption, $\exists \; u \in \mathcal{H}_u, \; y \in \mathcal{H}_y,$ such that

$$y_o = y - W u .$$

So
$$\hat{u} = -(R + W^* Q W)^{-1} W^* Q y + (R + W^* Q W)^{-1} W^* Q W u$$
$$= -[(R + W^* Q W)^{-1} W^*] Q y - [(R + W^* Q W)^{-1}] R u . \qquad 2.7.12$$

Therefore, $\hat{u}$ is bounded, since only bounded operators and vectors appear on the right of 2.7.12.

But
$$(R + W^* Q W) \hat{u} = -W^* Q y_o$$

i.e.
$$R \hat{u} + W^* Q \hat{y} = 0 .$$

<u>Example 2.7.1</u>: Consider the system

$$y(t) = e^{+t} + \int_o^t e^{(t - \tau)} u(\tau) \, d\tau$$

or
$$y(s) = \frac{1}{s - 1} (1 + u(s)) .$$

We wish to minimise

$$J = \int_0^\infty y^2 + u^2 \, dt \, .$$

For this system stabilising controls are known. In fact, if we propose

$$u(s) = - \frac{k}{s + a}$$

Then $$y(s) = \frac{1}{s - 1} \left(\frac{s + a - k}{s + a} \right)$$

which is stable for

$$a - k = - 1$$

and $$a > 0 \, .$$

i.e. $$k = a + 1$$

which gives $$u(s) = - \frac{(a + 1)}{s + a}$$

$$y(s) = \frac{1}{s + a} \, .$$

For optimality, we require

$$u(t) + \int_t^\infty e^{(\tau - t)} y(\tau) \, d\tau = 0 \, .$$

For $\ddot{u}(t)$ we propose the form

$$u(t) = - (a + 1) e^{-at}$$

Then $$- (a + 1) e^{-at} + \int_0^\infty e^{\xi} e^{-a\xi} e^{-at} d\xi = 0$$

$$(- (a + 1) + \int_0^\infty e^{(1-a)\xi} \, d\xi) e^{-at} = 0$$

which shows that the form of $u(t)$ is correct. If $a > 1$, the integral on the left converges, to give

$$-(a+1) + \frac{1}{1-a} = 0$$

i.e.
$$a = \sqrt{2}$$

and
$$u(t) = -\left(\frac{\sqrt{2}+1}{s+\sqrt{2}}\right) .$$

This example, though numerically trivial, presents a few important points. The first is that while all that is necessary for stability is $a > 0$, the gradient is infinite unless $a > 1$. So gradient type algorithms may not work for unstable systems. Even if g exists, finite, $\Delta u = -\varepsilon g$ may not be in the domain of W, and hence $u + \Delta u$ will not stabilise the system.

The spectral factorisation methods may still be applicable. The operator $R + W^* Q W$, though unbounded, may have a factorisation such that

$$F^* F = R + W^* Q W$$

where F is unbounded in general, but

$$\mathcal{D}(F) = \mathcal{D}(W)$$

and F^{-1} exists and is bounded. Then, as for the bounded case

$$\hat{u} = -F^{-1} [F^{*-1} W^* Q y_o]_+ .$$

If y_o is unbounded, then the projection operation gives an unbounded response, which when operated on by F^{-1} must be bounded. From the previous example

$$R + W^* Q W = 1 + \frac{1}{-s-1} \cdot \frac{1}{s+1}$$

$$= \frac{\sqrt{2}+s}{1-s} \cdot \frac{\sqrt{2}-s}{1+s}$$

$$= F \cdot F^* .$$

$$u(s) = -\frac{(1-s)}{\sqrt{2}+s} \left[\frac{1+s}{\sqrt{2}-s} \cdot \frac{1}{1+s} \cdot \frac{1}{1-s} \right]_{\substack{\text{poles of} \\ y_o(s) \text{ only}}}$$

$$= -\frac{(1-s)}{\sqrt{2}+s} \cdot \frac{1}{\sqrt{2}-1} \cdot \frac{1}{1-s}$$

$$= -\frac{\sqrt{2}+1}{\sqrt{2}+s} .$$

This method is not very practical when the weighting functions and disturbances are only known numerically. However, for the causal operators on discrete or continuous systems, though they may be unstable on the infinite time interval, on any finite interval $[0, T]$ they are bounded if their weighting functions are exponentially bounded, as is usually the case. Hence gradient algorithms can be used to solve finite time problems. However, if the optimal trajectories of both u and y converge to zero in this interval, it may be possible to infer that these u and y constitute a close approximation to the optimal solution. Although in a practical design such a u cannot be applied open loop, the compensator designed by the methods of Section 2.10 may stabilise the system, even for small inaccuracies in the optimal trajectories. In fact, most of the examples of Section 2.4 were unbounded on the infinite interval, due to the presence of pure integrators. The comments made in Section 2.4 about the choice of optimisation interval are particularly relevant, to avoid the ratio of minimum to maximum bounds on $R + W^* Q W$ becoming too small.

A further method of optimising unstable systems is first to stabilise the system by feedback, and then optimise as for bounded systems. Unfortunately, this leads to a cross-coupled cost function, but this is no essential difficulty.

The existence of stabilising controls is not a trivial problem. In the case of systems which have a finite state-space description, with initial conditions as disturbances, controllability is enough to guarantee the existence of stabilising controls. However, for infinite dimensional systems, the answer is not obvious. Consider the time invariant system whose impulse response is a square wave (Figure 2.7.1).

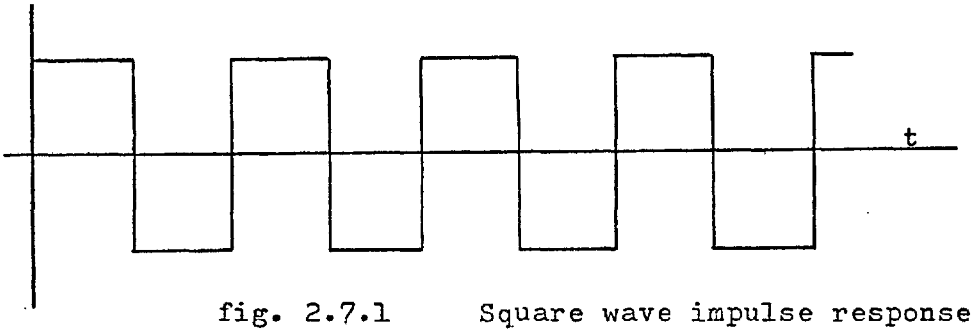


fig. 2.7.1 Square wave impulse response.

This may arise from an ideal model of a vibrating elastic bar. The Laplace transform $\frac{\tanh s}{s}$ has poles at $s = \pm (n + \frac{1}{2}) j \pi$, and the existence of stabilising controls is not obvious. One would not expect results obtained from truncation to be applicable here to the infinite time case.

2.8 State Space Solutions

Quadratic optimality for finite dimensional linear systems, described by state equations

$$\begin{aligned} \frac{dx(t)}{dt} &= A(t)\, x(t) + B(t)\, x(t) \\ y(t) &= C(t)\, x(t) \end{aligned} \qquad 2.8.1$$

for continuous systems, or

$$\begin{aligned} x_{k+1} &= A_k\, x_k + B_k\, u_k \\ y_k &= C_k\, x_k \end{aligned} \qquad 2.8.2$$

for discrete systems, has been extensively investigated in the literature. The state-vector $x(t)$ is an n-dimensional vector, y is p-dimensional, u is m-dimensional, and the matrices A, B, C, Q, R are of the appropriate size. The relevant cost functions

$$J = \int_{o}^{T} x^T Q x + u^T R u \, dt \qquad 2.8.3$$

or $$J = \sum_{k=0}^{N} (x^T Q x + u^T R u)_k \qquad 2.8.4$$

are to be minimised.

The solutions to this problem, derived via dynamic programming are well known. They rely on Bellman's principle of optimality which states that any portion of an optimal trajectory is an optimal trajectory. Using this principle, and the state concept, an optimal control law is obtained, which states that $u(t)$ is an instantaneous function of state and time.

For continuous systems, the optimum $u(t)$ is given by

$$u(t) = - R^{-1}(t) \, B^T(t) \, P(t) \, x(t) \qquad 2.8.5$$

where $P(t)$ is the $n \times n$ matrix given by the solution of the Riccati differential equation

$$- \frac{dP(t)}{dt} = Q + P A + A^T P - P B R^{-1} B^T P \qquad 2.8.6$$

$$P(t) = 0 \quad .$$

The optimal cost from time t to T is then

$$J(t) = x^T(t) \, P(t) \, x(t) \quad . \qquad 2.8.7$$

Similar results are obtained for sampled systems, where

$$u_k = (R_k + B_k^T P_{k+1} B_k)^{-1} B_u^T P_{k+1} A_k x_k \qquad 2.8.8$$

and

$$P_k = Q_k + A_k^T P_{k+1} - P_{k+1} B_k (R_k + B_k^T P_{k+1} B_k)^{-1} B_k^T . P_{k+1} A_k \qquad 2.8.9$$

$$P_N = Q_N$$

and also

$$J_k = x_k^T P_k x_k \quad . \qquad 2.8.10$$

For the continuous time case, these results are derived from the conditions of Section 2.2, to show the link between the methods. The fundamental property of the state-space approach is the semi-group law obeyed by the transition matrices

$$\Phi(t, \tau) = \Phi(t, s) \Phi(s, \tau) . \qquad 2.8.11$$

Now

$$\begin{aligned} x(t) &= \Phi(t, o) x_o + \int_o^t \Phi(t, \tau) B(\tau) u(\tau) d\tau \\ &= x_o(t) + \int_o^t W(t, \tau) u(\tau) d\tau \quad . \end{aligned} \qquad 2.8.12$$

Then for R P.D., Q P.S.D. (and the system controllable), the necessary and sufficient condition for optimality is

$$R(t) u(t) + \int_t^T W^T(\tau, t) Q(\tau) x(\tau) d\tau = 0 \quad . \qquad 2.8.13$$

However $$\int_t^T W^T(\tau, t)\, Q(\tau)\, x(\tau)\, d\tau$$

$$= \int_t^T B^T(t)\, \Phi^T(\tau, t)\, Q(\tau)\, x(\tau)\, d\tau$$

$$= B^T(t) \int_t^T \Phi^T(\tau, t)\, Q(\tau)\, x(\tau)\, d\tau$$

$$= B^T(t)\, \lambda(t)$$

where $$\lambda(t) = \int_t^T \Phi^T(\tau, t)\, Q(\tau)\, x(\tau)\, d\tau \quad . \qquad 2.8.14$$

Now, from theorem 2.2.3

$$\hat{J}(t) = \int_t^T x_t^T(\tau)\, Q\, x(\tau)\, d\tau \qquad 2.8.15$$

where $$x_t(\tau) = \Phi(\tau, t)\, x(t) \quad . \qquad 2.8.16$$

i.e. $$\hat{J}(t) = x^T(t)\, \lambda(t) \qquad 2.8.17$$

and $$u(t) = - R^{-1}(t)\, B^T(t)\, \lambda(t) \quad . \qquad 2.8.18$$

Now $$\lambda(t) = \int_t^T \Phi^T(\tau, t)\, Q(t)[\, \Phi(\tau, t)\, x(t) - \int_t^\tau \Phi(\tau, s)\, B(s)\, R^{-1}(s)\, B^T(s)\, \lambda(s)\, ds]\, d\tau$$

$$= [\int_t^T \Phi^T(\tau, t)\, Q(\tau) \Phi(\tau, t)\, d\tau] x(t) - \int_t^T \Phi^T(\tau, t)\, Q(t) \int_t^T \Phi(\tau, s)\, B(s)\, R^{-1}(s)\, B^T(s)\, \lambda(s)\, ds\, d\tau$$

This is a linear integral equation for $\lambda(s)$, $t \leqq s \leqq T$ and hence the solution is linearly dependent on $x(t)$. In particular

$$\lambda(t) = P(t)\, x(t) \quad . \qquad 2.8.19$$

Then
$$\begin{aligned} u(t) &= - R^{-1}(t)\, B^T(t)\, P(t)\, x(t) \\ &= - K(t)\, x(t) \end{aligned} \qquad 2.8.20$$

and
$$\hat{J}(t) = x^T(t)\, P(t)\, x(t) \quad . \qquad 2.8.21$$

But
$$\begin{aligned} \hat{J}(t) &= \int_t^T (x^T Q x + u^T R u) dt \\ &= \int_t^T (x^T Q x + x^T K^T R K x) dt \\ &= \int_t^T x^T [Q + K^T R K] x\, dt \quad . \end{aligned} \qquad 2.8.22$$

$$\begin{aligned} x(\tau) &= \Phi(\tau, t)\, x(t) + \int_t^\tau \Phi(\tau, s)\, B^T(s)\, K(s)\, x(s)\, ds \\ &= \Psi(\tau, t)\, x(t) \end{aligned} \qquad 2.8.23$$

on solving the resultant integral equation.

Then $$J(t) = \int_t^T x^T(t)\Psi^T(\tau, t)\,(Q + K^T R\,K)\Psi(\tau, t)\,x(t)\,d\tau$$

$$= x^T(t)\left[\int_t^T \Psi^T(\tau, t)\,(Q + K^T R\,K)\,\Psi(\tau, t)\,d\tau\right]x(t)$$

$$= x^T(t)\;P(t)\;x(t)\quad .$$

So $$P(t) = \int_t^T \Psi^T(\tau, t)\,[\,Q + K^T R\,K\,]\,\Psi(\tau, t)\,d\tau\;. \qquad 2.8.24$$

Therefore, without loss of generality, $P(t)$ can be symmetric, and also $P(T) = 0$.

Since $P(t)$ is expressed as an integral, it possesses a derivative a.e. Differentiate both sides with respect to t .

$$-\frac{dP}{dt} = Q(t) + K^T R\,K\,(t) - \int_t^T \frac{\partial\,\Psi^T(\tau,\,t)}{\partial t}\,[Q + K^T R\,K]\Psi(\tau,t)d\tau$$

$$-\int_t^T \Psi^T(\tau,\,t)[\,Q + K^T R\,K]\,\frac{\partial\Psi}{\partial t}(\tau,\,t)\,d\tau\quad .$$

Now $$x(\tau) = \Psi(\tau,\,t)\,x(t)\quad .$$

So $$\frac{\partial x(\tau)}{\partial t} = \frac{\partial\,\Psi(\tau,\,t)}{\partial t}\,.\,x(t) + \Psi(\tau\,,\,t)\,(A - B\,K)\,x(t)$$

$$= 0 \qquad \forall\; x\quad .$$

i.e. $$\frac{\partial\,\Psi(\tau,t)}{\partial t} = -\Psi(\tau,\,t)\,(A - B\,K)\;. \qquad 2.8.25$$

$$\text{So} \quad -\frac{dP}{dt} = Q + K^TRK + (A - BK)^TP + P(A - BK)$$

$$= Q + A^TP + PA + PBR^{-1}B^TP - PBR^{-1}B^TP - PBR^{-1}B^TP$$

$$= Q + A^TP + PA - PBR^{-1}B^TP \quad . \qquad 2.8.26$$

This derivation has been rather long winded compared to the simple dynamic programming argument, but the link between integral and differential methods is established. In our derivation, no use was made of the fact that $x(t)$ is a finite-dimensional vector. A similar procedure can be used for sampled-data systems. An alternative derivation has been given by Luenberger for the time-invariant case [L 4] .

2.9 Spectral Factorisation

An important operation to be performed in control calculations is the factorisation of a self-adjoint operator A into the product of a causal operator and its adjoint, or, more generally, the product of a causal operator, a self-adjoint instantaneous operator, and the adjoint of a causal operator. We shall assume that A has a bounded inverse. Then

$$A = F^* F \qquad 2.9.1$$

or

$$A = T^* S T = (I + G^*)S(I + G) \qquad 2.9.2$$

where
$$F, T \in \mathcal{J}_+$$
$$F^*, T^* \in \mathcal{J}_- \qquad 2.9.3$$
and
$$S \in \mathcal{J}_+ \cap \mathcal{J}_- \ .$$

A further restriction that is imposed is that the operators F, T and S all have bounded inverses, though these factorisations may not be unique. The operators F and T are bounded only if A is bounded. In this section, we first present general methods for the solution of the problem, followed by particular methods for special cases.

1. The solution of the optimal control problem for all

$$y_o = [W \alpha]_+$$

specifies G of 2.9.2 uniquely, where

$$\hat{u} = -(G \hat{u} + [G \alpha]_+) \ .$$

In control problems, when the P.D. operator $R + W^* Q W$ has a bounded inverse, we know that a factorisation exists, because of the existence and uniqueness of the optimal control.

2. Any method used to solve the control problem can be turned into a method of spectral factorisation. The iterative methods of Section 2.6 are particularly useful.

$$\begin{aligned}(I + G)^{-1} &= (I + G - G)(I + G)^{-1} \\ &= I - G(I + G)^{-1} \qquad 2.9.4 \\ &= I - (I + G)^{-1} G \ . \qquad 2.9.5\end{aligned}$$

Put

$$M = G(I + G)^{-1} \qquad 2.9.6$$

$$\in \mathcal{J}_+ \setminus \mathcal{J}_+ \cap \mathcal{J}_- \quad \text{if } G \text{ does.}$$

Also, if G is compact, so is M. The method below finds M rather than G. However, we can always find G from M, since

$$M + M\,G = G \qquad 2.9.7$$

and the iteration scheme

$$G_{k+1} = M + M\,G_k \qquad 2.9.8$$

is usually convergent since it is generally a functional equation of Volterra type.

Now

$$\begin{aligned}(I + G)^{*}S &= A(I - M) \\ &= A - A\,M \ . \qquad 2.9.9\end{aligned}$$

Now put

$$A = A_+ + A_- + D$$

as in 2.6.2-5. Then

$$S + G^{*} S = A_+ + A_- + D - A_+M - A_-M - D\,M \ . \qquad 2.9.10$$

Now project 2.6.10 into the various mutually exclusive subspaces.

Then
$$S = D \quad . \qquad 2.9.11$$

$$0 = A_+ - A_+M - D\,M - [A_-M]_+ \quad . \qquad 2.9.12$$

$$G^* S = A_- - [A_-M]_- \quad . \qquad 2.9.13$$

We concentrate on 2.9.12, which can be arranged into any iterative form desired. E.g.

(i) Jacobi method

$$M = D^{-1}(A_+ - A_+M - [A_-M]_+) \quad . \qquad 2.9.14$$

(ii) Relaxation method (Gauss-Seidel for $\omega = 1$; Jacobi for $\omega = 0$)

$$M = (D + \omega A_+)^{-1}(A_+ - (1-\omega)A_- - [A_-M]_+). \quad 2.9.15$$

Example 2.9.1 Laplace transform domain

Factorise
$$1 + \frac{1}{(1-s)(1+s)} \quad .$$

Answer
$$(1 + G) = \frac{\sqrt{2} + s}{1 + s}$$ (see example 2.5.1)

and
$$M = G(1 + G)^{-1}$$

$$= \frac{\sqrt{2} - 1}{\sqrt{2} + s} \quad .$$

However, using the Gauss-Seidel method, we obtain

$$1 + \frac{1}{(1-s)(1+s)} = 1 + \frac{\frac{1}{2}}{1-s} + \frac{\frac{1}{2}}{1+s} \quad .$$

$$(D + A_+)^{-1} = \left(1 + \frac{\frac{1}{2}}{1+s}\right)^{-1}$$

$$= \frac{2(1+s)}{2s+3} \quad .$$

$$M_{k+1}(s) = \frac{1+s}{2s+3}\left(\frac{1}{1+s} - \left[\frac{1}{1-s} M_k(s) \right]_+ \right).$$

$$M_o = 0 \qquad\qquad M_o(t)\big|_{t=0} = 0$$

$$M_1 = \frac{1}{2s+3} \qquad\qquad M_1(t)\big|_{t=0} = 0.5$$

$$M_2 = \frac{8s+13}{5(2s+3)^2} \qquad\qquad M_2(t)\big|_{t=0} = 0.4$$

$$M_3 = \frac{416s^2 + 1380s + 984}{125(2s+3)^3} \qquad\qquad M_3(t)\big|_{t=0} = 0.416$$

$$M_\infty(t)\big|_{t=0} = 0.414 \ldots$$

The main difficulty in the transform domain is the partial fraction separation. No such difficulty exists in the time domain. In fact, for more general continuous systems, specified by their impulse response, we find

$$R(t)\, u(t) + \int_o^\infty A(t, \tau)\, u(\tau)\, d\tau = - \int_t^\infty W^T(\tau, t)\, Q(\tau)\, y_o(\tau)\, d\tau$$

$$= f(t).$$

and $\int_0^\infty A(t, \tau)\, u(\tau)\, d\tau = \int_0^t B(t, \tau)\, u(\tau)\, d\tau + \int_t^\infty B(\tau, t)\, u(\tau)\, d\tau$.

We wish to find $M(t, \tau)$ where, equating the kernels of 2.9.14, we obtain

$$R(t)\, M(t, \tau) = B(t, \tau) - \int_\tau^t B(t, s)\, M(s, \tau)\, ds$$

$$- \int_t^\infty B(s, t)\, M(s, \tau)\, ds \; .$$

The procedure is very similar to the methods of Section 2.6.

3. The classical method of separating rational power spectral densities into reciprocal poles and zeros with respect to the $j\omega$ axis in the s domain, or the unit circle in the z domain is perfectly valid for simple analytical examples. However, it is not a computationally attractive method, since it is not easily mechanised. Youla [Y 1], Davis [D 1] and other writers have discussed these kinds of techniques for multivariable systems, i.e. square matrices whose elements are rational functions of s.

4. If the control system is amenable to a state-space treatment, then the spectral factorisation problem can be solved by means of the Matrix-Riccati equation. If $\hat{u}$, $\hat{x}$ are optimal trajectories, then

$$\hat{u} = - K \hat{x} \qquad 2.9.16$$

where

$$K \in \mathcal{J}_+ \cap \mathcal{J}_- \; . \qquad 2.9.17$$

Also
$$\hat{x} = \Phi(x_o \delta + B \hat{u}) \tag{2.9.18}$$

where δ is the impulse function, and Φ maps the space of state-trajectories into itself.

Then
$$\begin{aligned} \hat{x} &= \Phi x_o \delta + W \hat{u} \\ &= [W \alpha]_+ + W \hat{u} \end{aligned} \tag{2.9.19}$$

if all states are controllable from the input.

<u>Theorem 2.9.1:</u>

$$(I + B^T \Phi^* K^T) S (I + K \Phi B) = R + W^* Q W \tag{2.9.20}$$

where
$$S \in \mathcal{J}_+ \cap \mathcal{J}_- \quad .$$

<u>Proof</u>:
$$\begin{aligned} \hat{u} &= - K \Phi x_o \delta - K W \hat{u} \\ &= -[K \Phi B \alpha]_+ - (K \Phi B) \hat{u} \qquad (2.9.21) \\ &= -[G \alpha]_+ - G \hat{u} \quad . \end{aligned}$$

Hence by the results of Section 2.5, equations 2.5.23-40

$$(I + B^T \Phi^* K^T) S (I + K \Phi B) = R + W^* Q W \; .$$

For the particular case of continuous time-invariant dynamics, Q and R constant matrices, and infinite cost interval, these results can be stated in the frequency domain as a generalisation of a result of Kalman's [K 2] .

Theorem 2.9.2:

$$(I + B^T \Phi^T(-s)\, P\, B\, R^{-1}) R (I + R^{-1}\, B^T\, P\, \Phi(s)\, B)$$

$$= R + B^T \Phi^T(-s)\, Q\, \Phi(s)\, B \qquad 2.9.22$$

where $$Q + P A + A^T P - P B R^{-1} B^T P = 0 \quad , \qquad 2.9.23$$

P is positive definite, symmetric

and $$\Phi(s) = (s I - A)^{-1} \ . \qquad 2.9.24$$

Proof: Kalman has proved the result for single input, single output systems, with $R = 1$. The generalisation is very simple to prove and follows Kalman's proof very closely. We shall omit the proof. However, we shall prove a similar, but not so well known, result, for discrete time systems.

Theorem 2.9.3: For discrete time systems, under the same stationarity conditions as Theorem 2.9.2,

$$(I + B^T \Phi^T(\tfrac{1}{z})\, K^T)(R + B^T\, P\, B)(I + K^T \Phi(z)\, B)$$

$$= R + B^T \Phi^T(\tfrac{1}{z}) R \Phi(z)\, B \qquad 2.9.25$$

where $$P = Q + A^T P A + A^T P B K \qquad 2.9.26$$

with P positive definite, symmetric ,

$$K = (R + B^T\, P\, B)^{-1}\, B^T\, P\, A \qquad 2.9.27$$

and $$\Phi(z) = (z I - A)^{-1} \ . \qquad 2.9.28$$

Proof: From 2.9.26

$$P - A^T P A + A^T P B K = Q$$

$$z(A^T P - A^T P) + z^{-1}(P A - P A) + P - A^T P A + A^T P B K = Q.$$

$$(z^{-1} I - A^T)P(z I - A) + A^T P(z I - A) + (z^{-1} I - A^T)PA + A^T P B K = Q$$

Then multiply on the left by $\Phi^* = \Phi^T(\frac{1}{z})$, and on the right by Φ.

$$P + \Phi^* A^T P + P A \Phi + \Phi^* A^T P B K \Phi = \Phi^* Q \Phi .$$

Now multiply on the left by B^T, and on the right by B.

$$B^T P B + B^T \Phi^* A^T P B + B^T P A \Phi B + B^T \Phi^* A^T P B K \Phi B = B^T \Phi^* Q \Phi B$$

Now add R to both sides.

$$R + B^T P B + B^T \Phi^* A^T P B (R + B^T P B)^{-1}(R + B^T P B)$$

$$+ (R + B^T P B)(R + B^T P B)^{-1} B^T P A \Phi B + B^T \Phi^* A^T P B(R + B^T P B)^{-1}$$

$$(R + B^T P B)K \Phi B$$

$$= R + W^* Q W .$$

The left hand side can be simplified to give

$$(I + B^T \Phi^* K^T)(R + B^T P B)(I + K \Phi B) = R + W^* Q W .$$

It is known that $(I + K \Phi B)^{-1}$ is stable if P is the unique positive definite solution of 2.9.26. So the spectral factorisation problem is solved.

5. For discrete systems with signals in $l_2[0, N]$, $l_2[0, \infty)$ or more generally $R_m \otimes l_2[0, N]$ etc., a very simple numerical method is available for factorisation of self-adjoint operators, in terms of components. To be general, we describe the algorithm for operators mapping $R_m \otimes l_2[0, N]$ into itself.

If $$u \in R_m \otimes l_2[0, N] \quad , \qquad 2.9.29$$

then u can be represented as a sequence of m-vectors u_k. The self-adjoint operator A mapping u into b, i.e.

$$A\,u = b \qquad 2.9.30$$

is represented by

$$\sum_{k=0}^{N} A_{ik}\,u_k = b_i \qquad 2.9.31$$

where the A_{ik} are $m \times m$ matrices, and

$$A_{ik} = A_{ki}^T \,. \qquad 2.9.32$$

If u is represented as a column

$$u = \begin{pmatrix} \underset{\sim}{u}_1 \\ \underset{\sim}{u}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{u}_N \end{pmatrix} \quad , \qquad 2.9.33$$

A becomes a partitioned matrix of $m \times m$ blocks, i.e.

$$A = \left(\begin{array}{c|c|c|c} A_{11} & A_{12} & \cdots & A_{1N} \\ \hline A_{21} & A_{22} & & \\ \hline & & & \\ \hline A_{N1} & & & A_{NN} \end{array}\right) = A^T \qquad 2.9.34$$

Now this is required to be factorised into the product of a causal operator and its adjoint

i.e. $$F^* F = A . \qquad 2.9.35$$

However, a causal operator is represented by a lower block triangular matrix, since it can only operate on past and present time. Hence the factorisation problem reduces to decomposing a symmetric matrix into the product of an upper triangular matrix and its transpose. When A is positive definite, this decomposition is unique, and very easy to perform computationally via the Cholesky algorithm[P 1; p.8. 19] . Let a_{ij} be the elements of the $mN \times mN$ matrix A, and u_{ij} be the elements of F. Then the algorithm is organised to replace the lower half of A with F.

Algorithm 2.9.1:

```
For i = 1 to mN
    For j = 1 to i
        For r = 1 to j - 1
            $a_{ij} = a_{ij} - u_{ri} u_{rj}$
        repeat
            $u_{ij} = a_{ij}/u_{ii}$   if $i \neq j$
                 $= (a_{jj})^{\frac{1}{2}}$   $i = j$
    repeat
repeat .
```

This method is explicit and easy to perform computationally. Note Paige's comments on solution of the equation

$$A\,x = b$$

by first finding

$$F\,x = F^{x-1}\,b = c$$

and solving $x = F^{-1}\,c$.

This is precisely the factorisation method of optimal control solution that we propose in Section 2.5. Paige shows that this is an exceptionally stable numerical method, and is very insensitive to rounding errors. Note also that the solving of the equation

$$F\,x = c$$

is simply performed by back substitution, due to the triangularity of F.

The factorisation is only unique to within a unitary transformation. If U is any m x m unitary matrix, then

$$\bar{F} = \left(\begin{array}{c|c|c|c} U\,F_{11} & 0 & \cdot\cdot & 0 \\ \hline U\,F_{21} & U\,F_{22} & \cdot & 0 \\ \hline U\,F_{31} & U\,F_{32} & \ddots & \vdots \\ \hline \vdots & \vdots & \cdot & \vdots \end{array}\right) \tag{2.9.36}$$

will be another solution.

2.10 Feedback and Compensator Design

For a control system to be termed automatic, it must have built into it a method for recognising the presence of disturbances and correcting for them. This is most easily done by feedback. From classical design concepts it is known that well-designed feedback does more than realise particular controls for particular disturbances. Reduction of non-linear effects, disturbance rejection, and reduction of sensitivity to plant variations are all beneficial side effects. In fact, a feedback realisation of a control system is used because the knowledge of the plant and disturbance is deficient in some sense, and a basic problem is to design feedback systems which have good stability properties.

A general system is represented in Figure 2.10.1 in block diagram form. There may exist subsidiary outputs which are available for feedback, but not directly costed, and these are included in the y vector. The cost matrix Q will then be only semi-definite. H_1 and H_2 are general causal operators to be designed.

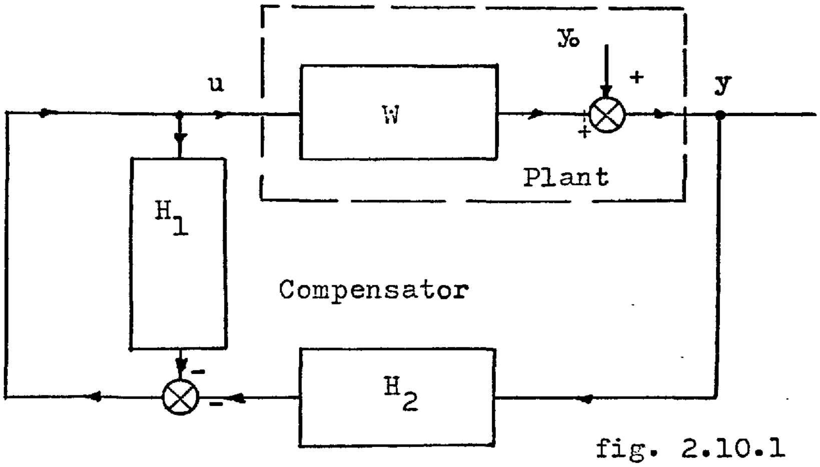


fig. 2.10.1

This configuration may allow the achievement of other design requirements besides a satisfactory performance index.

If W is the weighting function for a finite state system, y_o represents initial conditions, and y is a measurement of the complete state vector, then it is possible to minimise the performance index

$$J = \int_o^T x^T Q x + u^T R u \, dt$$

by setting $H_1 \equiv 0$, and H_2 the instantaneous operator $R^{-1}(t) \, B^T(t) \, P(t)$ using the notation of Section 2.8, and this is true for <u>all</u> initial conditions. In particular, if R, Q, A and B are time invariant, and $T = \infty$ then H is just a constant gain matrix, since $P(t) \longrightarrow P_\infty$, a constant positive definite matrix. If all the state vector is not available for feedback, it is not possible in general to achieve optimality for all initial conditions.

In this section, we present a method for achieving optimal control for particular disturbances, using a causal feedback compensator. We shall begin by assuming $H_1 \equiv 0$, and design a time invariant causal filter $H = H_2$ which allows optimal control for some initial disturbance. In fact, if u is an optimal control, and y is the corresponding optimal output, then

$$u = -H y \qquad 2.10.1$$

We particularise, and examine multivariable discrete and continuous time systems. In discrete time, 2.10.1 becomes

$$u_k = \sum_{i=0}^{k} H_{k-i}\, y_i \qquad 2.10.2$$

where the subscripts refer to time increments, u_k is m-dimensional, y_k is p-dimensional, and H_k is an $m \times p$ matrix. In order to solve for the H_k uniquely, we find p sets of independent optimal trajectories, using p sets of independent disturbances, and use the same H to relate these. Then, defining the matrices

$$\begin{aligned} Y_k &= (y^1,\ y^2,\ \ldots\ y^p)_k \\ U_k &= (u^1,\ u^2,\ \ldots\ u^p)_k \end{aligned} \qquad 2.10.3$$

one obtains

$$U_k = \sum_{i=0}^{k} H_{k-i}\, Y_i \ . \qquad 2.10.4$$

Providing Y_o is invertible, H_j is now uniquely specified. If $W_o = 0$, as is usual,

$$Y_o = Y_{o_o} \qquad 2.10.5$$

and hence Y_o is invertible if Y_{o_o} is. Then we obtain

Algorithm 2.10.1

$$\begin{aligned} -H_o &= U_o\, Y_o^{-1} \\ -H_1 &= (U_1 + H_o Y_1)\, Y_o^{-1} \qquad 2.10.6 \\ -H_k &= U_k + (H_o Y_k + H_1 Y_{k-1} + \ldots + H_{k-1} Y_1)\ \ Y_o^{-1} \end{aligned}$$

(simple back substitution). //

It is desirable that the sequence H_j is stable

i.e. $H_j \longrightarrow 0$ as $j \longrightarrow \infty$.

However, it is difficult to give necessary and sufficient conditions which ensure this. If the sequences U and Y are expressed in terms of z transforms, then

$$H(z) = - U(z) Y^{-1}(z) . \qquad 2.10.7$$

Since optimal controls are stable, U(z) is a stable transfer matrix. However, there is no guarantee that Y(z) is minimum phase. To obtain a stable H, $Y_o(z)$ should be chosen to make Y(z) M.P.

But $$Y = Y_o + W U \qquad 2.10.8$$

$$= Y_o - W H Y . \qquad 2.10.9$$

i.e. $$(I + W H)Y = Y_o \qquad 2.10.10$$

$$(I + W H)^{-1} = Y Y_o^{-1} . \qquad 2.10.11$$

But the closed loop system is stable, since both U and Y are stable. Hence $Y Y_o^{-1}$ will be stable, and so Y must be N.M.P. if Y_o is. However, Y_o minimum phase does not seem to guarantee that Y is also.

Examples 2.10.1: Figures 2.10.2-3 present the results of applying algorithm 2.10.1 to some of the examples of Section 2.4.

The same procedures are conceptually the same for continuous systems, but their numerical application is more difficult. We require to find K_i and $F(t)$ such that

$$- U(t) = \sum_i K_i Y(t - \tau_i) + \int_0^t F(t - \tau) Y(\tau) d\tau \qquad 2.10.12$$

where F is integrable. For constant R, the optimal $U(t)$ is absolutely continuous (by corollary to equation 2.4.5), except possibly at $t = 0$. Hence, $K_i = 0 \quad \forall i > 0$; i.e. without loss of generality

$$- U(t) = K_o Y(t) + \int_0^t F(t - \tau) Y(\tau) d\tau \quad . \qquad 2.10.13$$

We assume that $Y_o(0) = Y(0)$ is invertible. Then

$$- K_o = U(0) Y_o^{-1} \quad . \qquad 2.10.14$$

$$U(t) - U(0) = - \int_0^t F(\tau) Y(t - \tau) d\tau \quad . \qquad 2.10.15$$

To find F, we propose four methods.

1. If it is possible to take Laplace transforms of both sides of 2.10.15, then

$$- F(s) = (U(s) - U(0)) Y^{-1}(s) \qquad 2.10.16$$

Examples 2.10.1

1. System of example 2.4.2.   r= 0.1

Filter designed by back substitution algorithm.

```
      -0.14305E 01  0.95890E 00 -0.28992E 00  0.87649E-01
H
 k    -0.26499E-01  0.80000E-02 -0.24085E-02  0.71039E-03
      -0.18592E-03  0.74275E-04
```

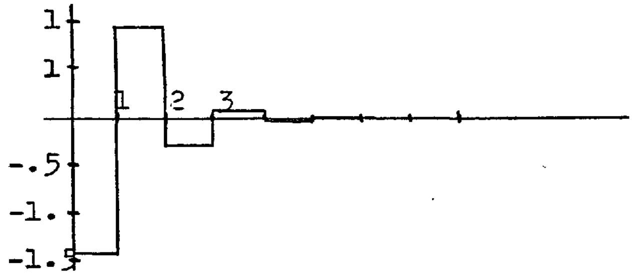


fig. 2.10.2

2. System of example 2.4.3   r= 0.1

fig. 2.10.3

```
     -0.12041E 01  0.74512E 00 -0.26531E 00  0.14393E 00
H    -0.89314E-01  0.57319E-01 -0.36127E-01  0.21027E-01
 11  -0.13095E-01  0.91930E-02 -0.55006E-02  0.28918E-02
     -0.22181E-02  0.13585E-02 -0.65352E-03  0.76591E-03
     -0.21577E-03 -0.21007E-03 -0.81408E-03 -0.50510E-03
```

```
      0.28268E 00 -0.76382E 00  0.69980E 00 -0.45767E 00
      0.28535E 00 -0.17777E 00  0.11140E 00 -0.69809E-01
H     0.43208E-01 -0.27307E-01  0.17054E-01 -0.10191E-01
 12   0.65146E-02 -0.47512E-02  0.21526E-02 -0.13660E-02
      0.13035E-02 -0.94048E-03 -0.29661E-03 -0.63880E-03
```

```
     -0.11647E 00 -0.66026E-01  0.10338E 00 -0.71329E-01
      0.43896E-01 -0.29286E-01  0.19097E-01 -0.11512E-01
H     0.63861E-02 -0.45056E-02  0.27301E-02 -0.14602E-02
 21   0.14581E-02 -0.89996E-03 -0.40704E-03 -0.57708E-03
      0.73727E-03  0.45490E-03 -0.17108E-04  0.50181E-04
```

```
     -0.91173E 00  0.74280E 00 -0.39420E 00  0.23836E 00
     -0.14688E 00  0.90658E-01 -0.57253E-01  0.36129E-01
H    -0.21872E-01  0.13595E-01 -0.91723E-02  0.57352E-02
 22  -0.28551E-02  0.18567E-02 -0.19486E-02  0.78868E-03
     -0.15866E-03  0.22707E-03 -0.70716E-03 -0.46938E-03
```

2. In the time domain, differentiate both sides of 2.10.15

$$- \frac{dU}{dt} = F(t)\,Y(0) + \int_0^t F(\tau)\,\frac{dY(t-\tau)}{d(t-\tau)}\,d\tau \quad . \qquad 2.10.17$$

$$- F(t) = \frac{dU}{dt}\,.\,Y(0)^{-1} + \left(\int_0^t F(t-\tau)\,\frac{dY(\tau)}{d\,\tau}\,d\tau\right)Y_0^{-1} \,. \qquad 2.10.18$$

Since 2.10.18 is a Volterra equation, this will converge from any starting point by simple successive substitution.

3. Neither method 1 nor 2 is particularly suitable if U and Y are only known numerically. A third method of solution is to place

$$E(t) = U(t) - U(0) + \int_0^t F(t-\tau)\,Y(\tau)\,d\tau \qquad 2.10.19$$

and minimise tr. $\int_0^\infty E^2(t)\,dt$ with respect to F, with such algorithms as the conjugate gradient method of section 2.3.

4. A method which can be used in conjunction with method 3 is to approximate the integral by a numerical integration procedure, and solve the resultant equations in a similar way to the discrete time case. This approximation is then refined using method 3.

The same stability problems arise with continuous time systems as with discrete systems. If

$$H(s) = K_0 + F(s) \qquad 2.10.20$$

then the equations 2.10.8-11 also hold for continuous time. Having calculated a feedback operator H by the methods above, one can use this directly as a feedback compensator, if it is suitable. However, it may be computationally convenient to use the more general configuration 2.10.1, where $H_1 \neq 0$.

Then, to achieve the same trajectories

$$H = (I + H_1)^{-1} H_2 \qquad 2.10.21$$

$$\therefore \quad H + H_1 H = H_2 \quad . \qquad 2.10.22$$

Considering the continuous time case as an example, it seems preferable that the weighting functions $H_1(t)$ and $H_2(t)$ decay as quickly as possible, for ease of simulation. Let H_2 have the same instantaneous transmission as H, i.e.

$$\begin{aligned} H &= K_o + F \\ H_2 &= K_o + F_2 \end{aligned} \qquad 2.10.23$$

$$F + H_1 K_o + H_1 F = F_2 \quad . \qquad 2.10.24$$

Using this as a dynamic equation for F_2 in terms of F_1, one possible design method is to minimise (for single input/output systems)

$$J = \int_o^\infty H_1^2 + \alpha H_2^2 \, dt \ . \qquad 2.10.25$$

The following property is a consequence of these methods:

Theorem 2.10.1: The compensated system, as desigred above is optimal for the entire subspace spanned by the initial disturbances which make up the columns of the matrix Y_o.

Proof: $$(R + W^* Q W) u^i = - W^* Q y_o^i$$

for each i, if u^i is optimal. Consider a disturbance

$$y_o = \sum_i \alpha_i y_o^i$$

where the α_i are scalars. Since the closed loop control system as designed above is linear, the control which will result is

$$u = \sum_i \alpha_i u_i .$$

But $$(R + W^* Q W) u = \sum_i \alpha_i (R + W^* Q W) u_i$$

$$= - \sum_i \alpha_i W^* Q y_o^i$$

$$= - W^* Q(\sum_i \alpha_i y_o^i)$$

$$= - W^* Q y_o .$$

i.e. the control is optimal for any linear combination of the y_o^i.

Example 2.10.2

Consider the continuous system represented by

$$y(s) = \frac{as + b}{s^2} + \frac{1 + s}{s^2} u(s) \ .$$

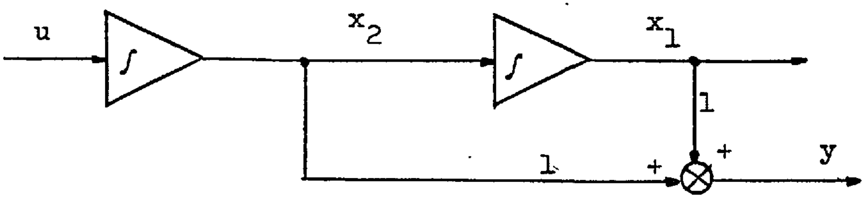


fig. 2.10.4

If we minimise

$$J = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} |y|^2 + |u|^2 \, ds$$

by the method of Section 2.5, we obtain

$$u(s) = - \frac{as + (\sqrt{3} - 1)a + (2 - \sqrt{3})b}{s^2 + \sqrt{3}s + 1}$$

$$y(s) = \frac{(\sqrt{3}b - b + a)s + b}{s^2 + \sqrt{3}s + 1} \ .$$

By the state space methods, one obtains

$$P = \begin{pmatrix} \sqrt{3} & 1 - \sqrt{3} \\ \sqrt{3} - 2 & 2\sqrt{3} - 3 \end{pmatrix}$$

and the optimal gains shown in Figure 2.10.5

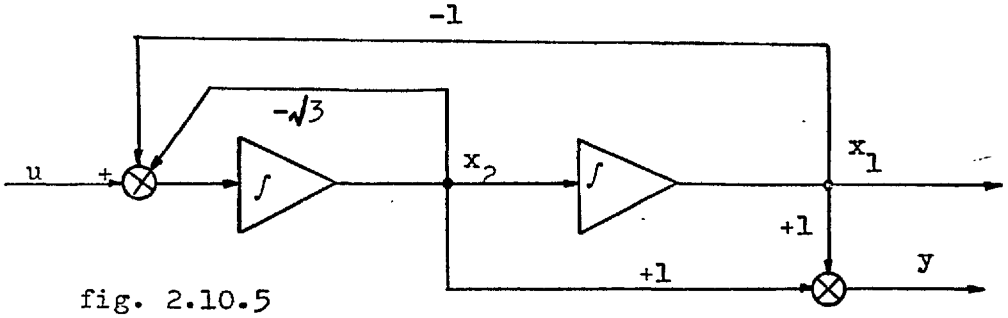


fig. 2.10.5

This also leads to the optimal $u(s)$, $y(s)$ as above. This configuration is optimal for all initial conditions. However, if only y is accessible, then $H(s)$ becomes

$$H(s) = -\frac{u(s)}{y(s)} = \frac{as + (\sqrt{3}-1)a + (2-\sqrt{3})b}{(\sqrt{3}b - b + a)s + b}$$

Note that this filter depends on the particular initial conditions. For $a = b = 1$, the disturbance y_o is just the impulse response of the system, and in this case we have the theorem:

<u>Theorem 2.10.2</u>: In a SI/SO system, if the initial condition y_o is the system impulse response W, then

$$G = H\,W \qquad 2.10.26$$

where $(1 + G^*)(1 + G) = 1 + W^* W$

and $H(s) = -\dfrac{\hat{u}(s)}{\hat{y}(s)}$.

<u>Proof</u>: When y_o represents initial conditions

$$\hat{u} = -G\hat{u} - [G\alpha]_+$$

from equation 2,5.38. Now here $\alpha = \delta(t)$ the unit impulse, so

$$\hat{u} = - G\,\hat{u} - G \qquad 2.10.27$$

and

$$\hat{y} = W(1 + \hat{u}) \;. \qquad 2.10.28$$

But

$$\begin{aligned} 1 + \hat{u} &= 1 - (1 + G)^{-1}\,G \\ &= (1 + G)^{-1} \;. \qquad 2.10.29 \end{aligned}$$

So

$$\hat{y} = W(1 + G)^{-1} \;.$$

But

$$\begin{aligned} H\,\hat{y} &= H\,W(1 + G)^{-1} = - \hat{u} \\ &= + G(1 + G)^{-1} \;. \end{aligned}$$

i.e.

$$G = H\,W \;.$$

This is true for our example. If $y_o \neq W(t)$ is non-minimum phase, then $H(s)$ may be unstable.

An alternative design procedure for designing a compensator is to try and build a system such that the total open loop transfer function is G. This will be optimal when $y_o(t) = W(t)$, for SI/SO systems. It is shown in the next section that this kind of system has excellent stability properties. One possible structure is shown in Figure 2.10.6

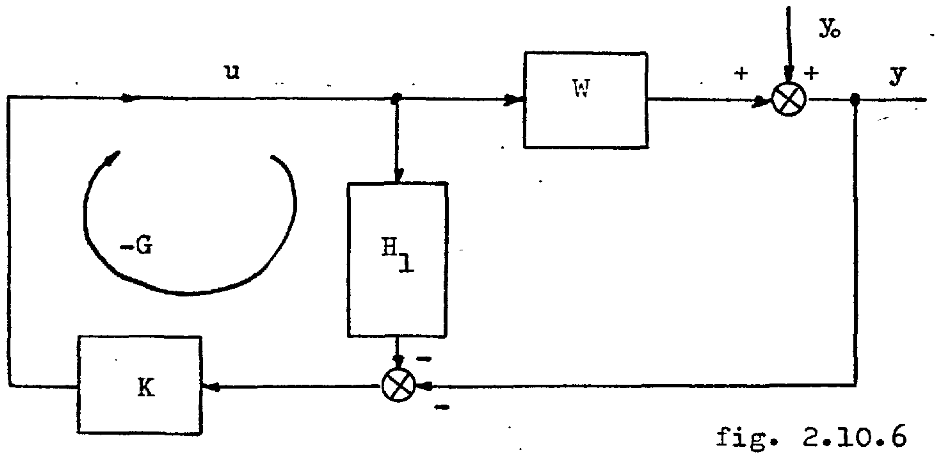


fig. 2.10.6

A more general type of structure which takes into account the disturbance y_o is developed in Chapter 3. However, the structure of Figure 2.10.6 may be useful.

Example 2.10.3
$$W(s) = \frac{e^{-s}}{s}$$

$$(1 + G(-s))(1 + G(s)) = 1 - \frac{1}{s^2}$$

$$G(s) = +\frac{1}{s}$$

Require
$$H_1 + W = \frac{1}{s} .$$

Then
$$H_1 = \frac{1 - e^{-s}}{s} .$$

$H_1(s)$ is the transfer function of a zero-order data hold with impulse response $H_1(t)$ as in Figure 2.10.7.

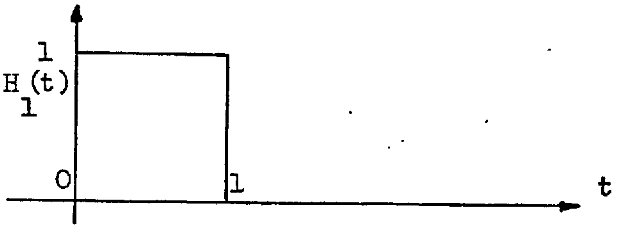


fig. 2.10.7

The overall control system becomes :

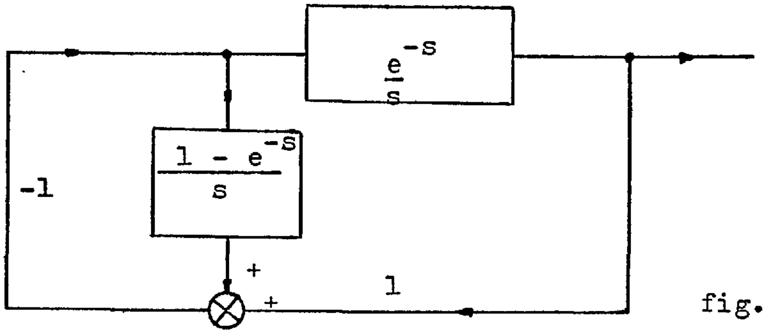


fig. 2.10.8

2.11 Properties of Optimal Systems

This section devotes its attention mainly to time-invariant single-input single-output continuous systems, which are specified in terms of their impulse response, or transfer function in the frequency domain. Then by optimality of a feedback system we shall mean a system with open loop gain G, where

$$|1 + G|^2 = 1 + |W|^2 \qquad 2.11.1$$

and 1 + G is minimum phase.

On the Nyquist disgram

$$|1 + G| > 1 \qquad 2.11.2$$

implies that 1 + G lies outside the unit circle. If G is plotted on the Nyquist diagram (Figure 2.11.1)

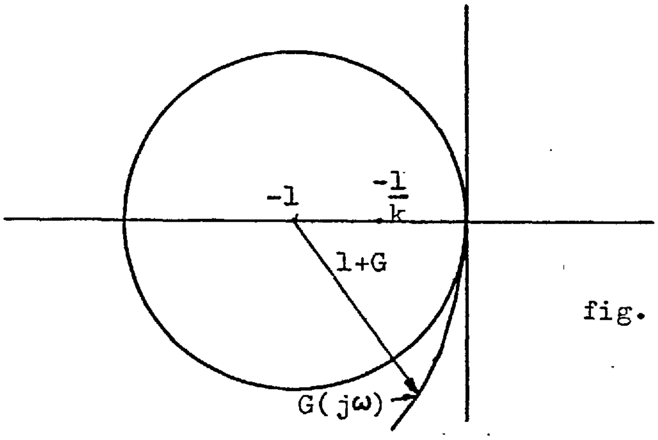


fig. 2.11.1

then G lies outside the unit circle centred on the -1 point. Since $(1 + G)^{-1}$ is stable, G must satisfy the Nyquist criterion for stability with gain 1. Then also:

<u>Theorem 2.11.1</u>: If G is an optimal transfer function, then $(1 + kG)^{-1}$ is stable for $k \in (\frac{1}{2}, \infty)$.

<u>Proof</u>: Since G lies entirely outside the unit circle centred on -1, it encircles the $-\frac{1}{k}$ point exactly the same number of times as it encircles the -1 point, where $-\frac{1}{k} \in (-2, 0)$. Hence $(1 + kG)^{-1}$ is stable for $k \in (\frac{1}{2}, \infty)$, since $(1 + G)^{-1}$ is stable.

<u>Corollary 2.11.1</u>: G has infinite gain margin.

<u>Proof</u>: $(1 + kG)^{-1}$ is stable as $k \longrightarrow \infty$.

<u>Corollary 2.11.2</u>: G is minimum phase.

<u>Proof</u>: If G is minimum phase, then, from the basic definition in Section 1.8, $\exists$ an F with infinite gain margin, such that

$$G\, F^{-1}(\varepsilon) \longrightarrow I \qquad \text{as } \varepsilon \longrightarrow 0$$

and for $\varepsilon < \varepsilon_{max}$, $F^{-1}(\varepsilon)$ is stable. But by corollary 2.11.1, $\varepsilon I + G$ has this property.

i.e. $(\frac{1}{k} + G)^{-1}$ is stable for $\frac{1}{k} < 2$.

$$\therefore \qquad G(\frac{1}{k} + G)^{-1} = (\frac{1}{k} - \frac{1}{k} + G)(\frac{1}{k} + G)^{-1}$$

$$= 1 - \frac{1}{k}(\frac{1}{k} + G)^{-1}$$

$$\to 1 \qquad \text{as} \quad k \longrightarrow \infty .$$

<u>Corollary 2.11.3</u>: $G(j\omega)$ has a phase margin greater than 60°.

<u>Proof</u>: The phase margin of a transfer function $G(j\omega)$ is the angle $180^\circ - \underline{/G(j\omega)}$, when $|G(j\omega)| = 1$. Consider Figure 2.11.2.

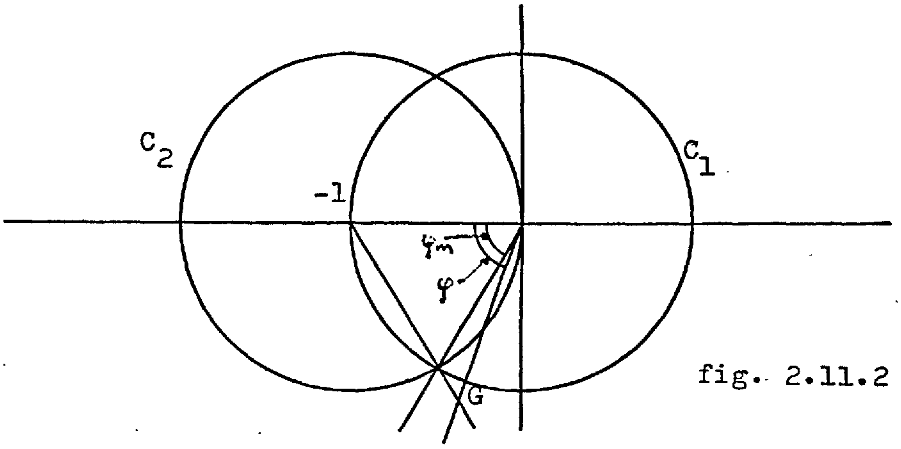


fig. 2.11.2

$G(j\omega)$ lies on the unit circle C_1 if $|G| = 1$. However, $G(j\omega)$ must lie outside circle C_2. The phase margin of G is the angle φ. But by construction $\varphi \geqq \varphi_m$. However, φ_m is the angle of an equilateral triangle.

$$\therefore \qquad \varphi \geqq 60^\circ \qquad\qquad 2.11.3$$

The phase margin is an indication of the damping and overshoot of the closed loop transient response.

Any transfer function G that possesses the properties

$$| 1 + G | > 1$$

and $(1 + G)^{-1}$ is stable, will be called optimal.

Theorem 2.11.2: If G is optimal, so is kG, for $k \geqq 1$.

Proof: $(1 + kG)^{-1}$ is stable for $k \geqq 1$ from Theorem 2.11.1.

But
$$
\begin{aligned}
| 1 + kG |^2 &= 1 + kG^* + kG + k^2G^*G \\
&= 1 + (k - 1)G^* + (k - 1)G + (k^2 - 1)G^*G + G^* + G + G^*G \\
&= 1 + |W|^2 + (k - 1)[G + G^* + G^*G] + (k^2 - k)G^*G \\
&= 1 + |W|^2 + (k - 1) | W|^2 + k(k - 1) | G |^2 \\
&> 1 \qquad \text{for } k \geqq 1.
\end{aligned}
$$

Examples 2.11.1, 2: The transcendental transfer functions

1. $$\frac{1}{\sqrt{s}}$$

2. $$\frac{1}{s + 0.5 + 0.8e^{-0.7s}}$$

are both optimal. (See Figure 2.11.3.)

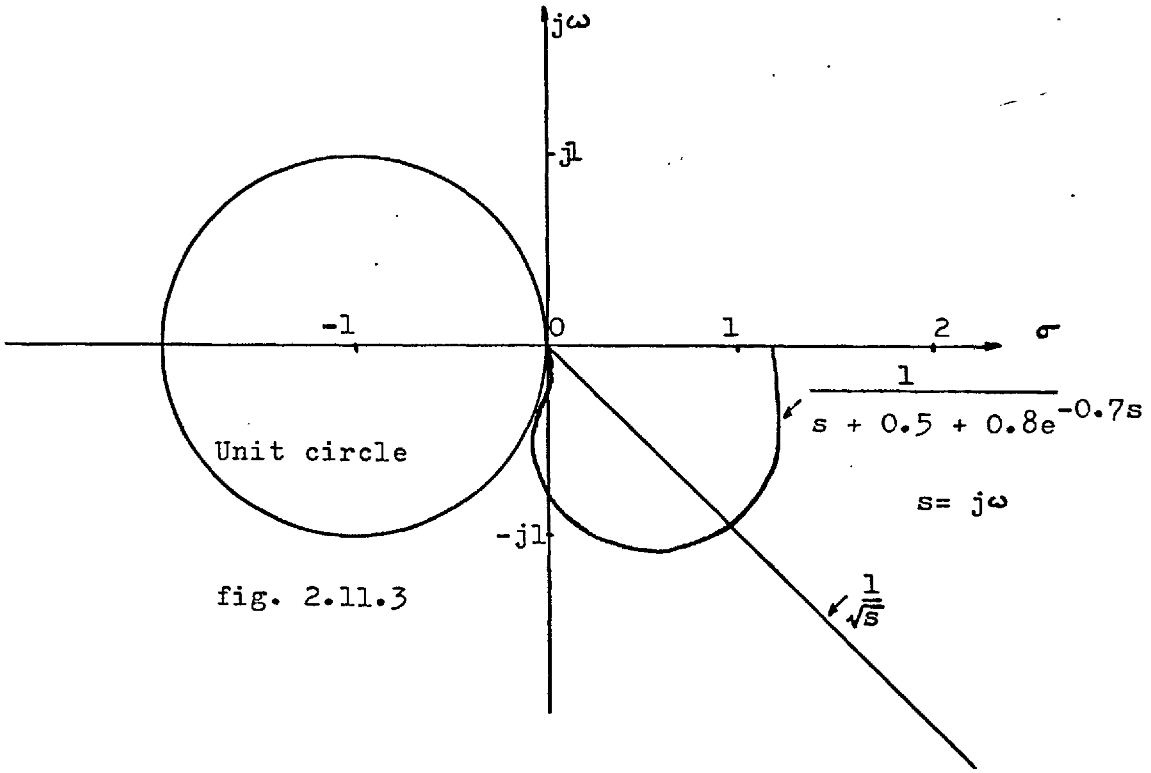


fig. 2.11.3

Lemma 2.11.1: If $g(t)|_{t=0}$ is finite, non zero, then

$$\lim_{\omega \to \infty} \underline{/G(j\omega)} = -90^{\circ} \qquad \text{if } G(s) \text{ is M.P.}$$

Proof: By initial value theorem

$$\lim_{|s| \to \infty} s\,G(s) = g(0) \qquad \text{under quite general conditions.}$$

But $$s\,G(s) = \text{Re } s\,G(s) + j.\,\text{Im } s\,G(s)\ .$$

Hence $$\lim_{|s| \to \infty} s\,G(s) = \lim_{|s| \to \infty} \text{Re } s\,G(s) + j \lim_{|s| \to \infty} \text{Im } sG(s)$$

$$= g(0) \neq 0 \text{ real, finite.}$$

i.e. $$\lim_{|s| \to \infty} \text{Im } sG(s) = 0$$

$$\lim_{|s| \to \infty} \angle s\, G(s) = \lim_{|s| \to \infty} \tan^{-1} \frac{\text{Im } s\, G(s)}{\text{Re } s\, G(s)}$$

$$= 0^\circ \quad , \quad \text{the minimum phase.}$$

But $$\angle s\, G(s) = \angle s + \angle G(s) \quad .$$

$$\therefore \quad \lim_{s \to \infty} \angle G(s) = - \lim_{s \to \infty} \angle s$$

$$= - 90^\circ.$$

A more comprehensive stability theorem than Theorem 2.11.1 can be proved, with the aid of the following geometrical lemma.

Lemma 2.11.2: If a point P lies outside a circle, then the angle subtended by the diameter of the circle at P is less than 90°.

Proof:

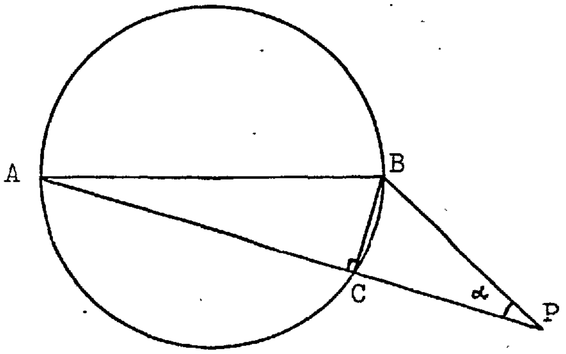


fig. 2.11.4

Consider the diagram of Figure 2.11.4, where AB is a diameter. At least one line, AP or BP, intersects the circle at C. But $\angle BCP$ is a right angle, since ACB lies in a semi-circle. But $\angle CBP \leqq 0$, and since the sum of the angles of a triangle is 180°

$$\angle CPB = \alpha \leqq 90^{\circ} . \qquad 2.11.4$$

In fact $\alpha = 90^{\circ}$ is only obtained when P lies on the circle.

Theorem 2.11.3: If $G(s)$ is an optimal open-loop transfer function, i.e. $|1 + G|^2 = 1 + |W|^2$, and $(1 + G)^{-1}$ stable, then the closed loop system in Figure 2.11.5

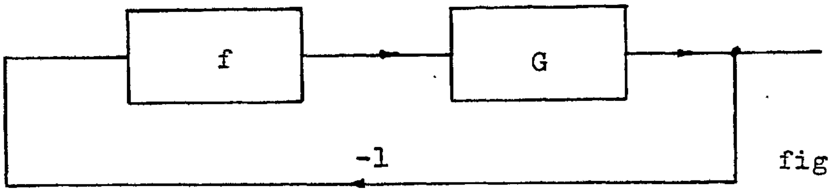


fig. 2.11.5

is asymptotically stable (in the sense of the Popov criterion) if f is a memoryless non-linearity, in the sector $(\frac{1}{2}, \infty)$ and obeys the assumptions of the Popov criterion.

Proof: The Popov criterion as formulated by Dewey [D 1] or Aizerman and Gantmacher [AG 1] gives a condition for stability with f in a sector $(0, k)$. So the given system must be transformed into such a system as shown in Figure 2.11.6.

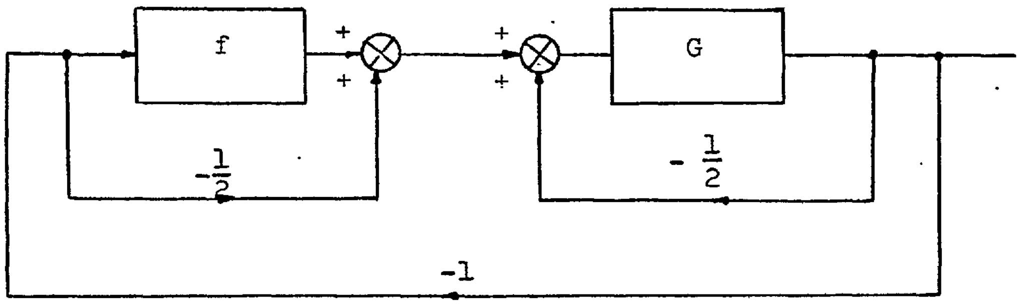


fig. 2.11.6

Now put $$\varphi = f - \tfrac{1}{2} \quad , \qquad 2.11.5$$

$$F = \frac{G}{1 + \frac{1}{2} G} \quad . \qquad 2.11.6$$

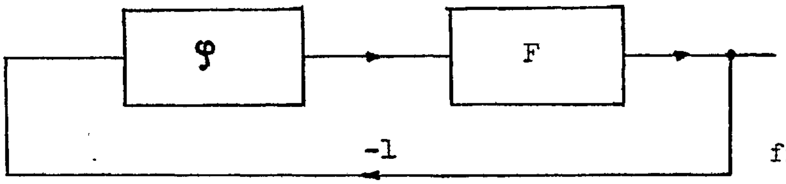


fig. 2.11.7

To satisfy the Popov criterion, we plot

$$F^{*}(j\omega) = \text{Re } F(j\omega) + j\omega \text{ Im } F(j\omega) \qquad 2.11.7$$

and the graph of $F^{*}(j\omega)$ must lie entirely to the right of a line intersecting the real axis at $-\frac{1}{k}$, for the system of Figure 2.11.7 to be stable with φ in sector $(0, k)$. However in this case

$$\text{Re } F^{*}(j\omega) = \text{Re } F(j\omega) \geqq 0. \qquad 2.11.8$$

For $$\text{Re } F(0) = \frac{G(0)}{1 + \frac{1}{2} G(0)} > 0 \quad \text{by assumption}$$

and $$| \underline{/F(j\omega)} | = | \underline{/G(j\omega)} - \underline{/2 + G(j\omega)} | \quad . \qquad 2.11.9$$

But $G(j\omega)$ lies entirely outside the circle of diameter 2 centred on the -1 point. (Figure 2.11.8).

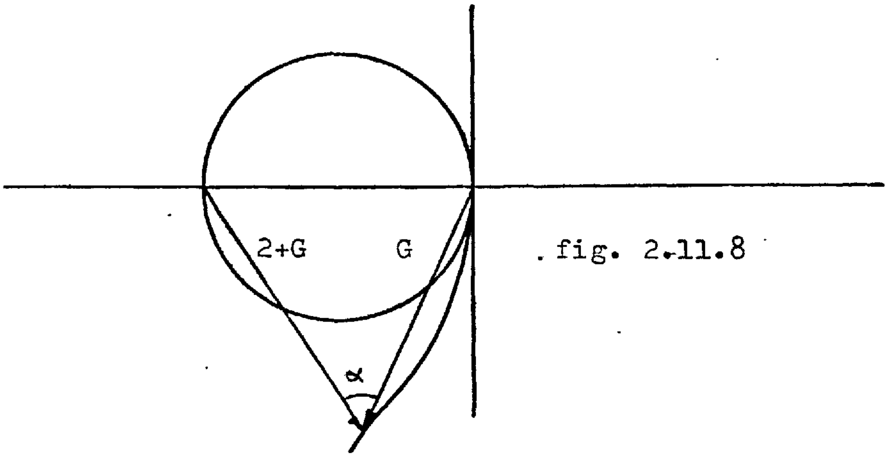


fig. 2.11.8

Hence, from lemma 2.11.2

$$\underline{/F(j\omega)} = \alpha \leqq 90^{\circ}$$

and so $$\text{Re } F(j\omega) \geqq 0 \qquad \forall \ \omega \ . \qquad 2.11.10$$

Hence, F and F^* lie entirely in the right half plane, and any vertical line in the left half plane is a Popov line, and so the system Figure 2.11.7 is stable for φ in the sector $(0, \infty)$; i.e. the original system in Figure 2.11.5 is stable for f in the sector $(\frac{1}{2}, \infty)$.

Kalman [K 2] investigated the asymptotic characteristics of the closed loop system, when

$$|1 + G|^2 = 1 + \rho|W|^2 \qquad 2.11.11$$

and $\rho \to \infty$. His conclusions are also shown to be true for non-rational G, W. The approach that we adopt uses the Bode diagram, rather than root-locus techniques which become inconvenient for non-rational functions. From 2.11.11

$$\frac{1}{|1 + G|^2} = \frac{1}{1 + \rho|W|^2} \qquad 2.11.12$$

$$\therefore \qquad \frac{\rho|W|^2}{|1 + G|^2} = \frac{\rho|W|^2}{1 + \rho|W|^2} .$$

$$= \frac{1}{1 + \frac{1}{\rho|W|^2}} \qquad 2.11.13$$

Now consider the behaviour of $\frac{\rho|W|^2}{|1 + G|^2}$ in various frequency ranges. (Note that $\frac{\rho^{\frac{1}{2}} W}{1 + G}$ is the closed loop transfer function from input to costed output.) For most systems, W is essentially low-pass, with maximum $|W|$ near zero frequency, and as frequency increases $|W(j\omega)|$ is asymptotic to $k\,\omega^{-\nu}$, where ν is real in general, and may be infinite. Define ω_o as the bandwidth of the closed loop system. I.e., for $\omega = \omega_o$

$$\rho|W|^2 = 1 .$$

$$2.11.14$$

Then $$\left.\frac{\rho|W|^2}{|1 + G|^2}\right|_{\omega=\omega_o} = \frac{1}{2} .$$

But if ρ is large enough, then for ω near ω_o

$$|W| \sim k\,\omega^{-\nu} . \qquad 2.11.15$$

i.e. $$\rho k^2 \omega_o^{-2\nu} = 1 .$$

$$\omega_o \sim (\rho k^2)^{\frac{1}{2\nu}} . \qquad 2.11.16$$

For $\omega << \omega_o$

$$\frac{\rho |W|^2}{|1 + G|^2} \simeq 1 .$$

For $\omega = \omega_o$

$$\frac{\rho^{\frac{1}{2}} |W|}{|1 + G|} = 0.707 .$$

Hence $$\frac{\rho |W|^2}{|1 + G|^2} \rightarrow \frac{1}{1 + \frac{\omega^{2\nu}}{\omega_o^{2\nu}}}$$

i.e. $$\frac{\rho^{\frac{1}{2}} |W|}{|1 + G|} \rightarrow \frac{1}{\sqrt{1 + (\frac{\omega}{\omega_o})^{2\nu}}} . \qquad 2.11.17$$

This behaviour is plotted in Figure 2.11.9.

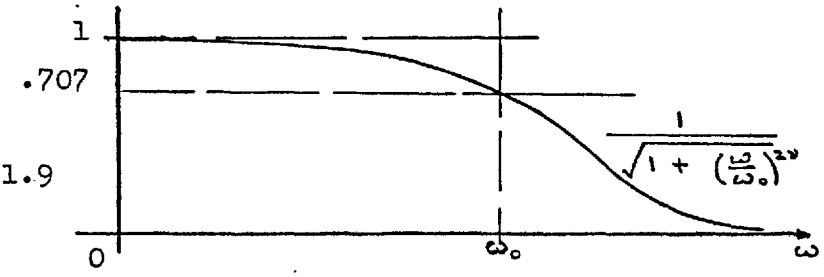


fig. 2.11.9

With reference to Guillemin [G 1], it is seen that this behaviour corresponds to the gain response of a Butterworth filter, for integral (finite) ν, and ν corresponds to the excess of poles over zeros of W. However, if this response characteristic is defined as generalised Butterworth for ν real, then Kalman's result can be generalised.

For W minimum phase, $\frac{W}{1+G}$ approaches a true Butterworth filter of order ν. However, if W is N.M.P., $W = W_1 A$, where W_1 is M.P. and A is all pass. $\frac{W}{1+G}$ is then asymptotic to the ν:th order Butterworth filter cascaded with the all-pass system A.

When $\nu \longrightarrow \infty$, the response in Figure 2.11.9 approaches that of the physically unrealisable ideal low pass filter. However, in this case, there will be no finite ω_o, since in general, we would need ρ' infinite.

i.e. $$\rho = \left(\frac{\omega_o}{k}\right)^{2\nu} \longrightarrow \infty \qquad \text{if} \quad \left(\frac{\omega_o}{k}\right) > 1.$$

The inverse Nyquist diagram is often used for control systems design. Circles in the Nyquist diagram, centred on the real axis, and passing through the origin will map into straight lines parallel to the imaginary axis, on the inverse Nyquist plot. In particular, a circle of radius a maps into a line distance $\frac{1}{2a}$ from the origin, as in Figure 2.11.10.

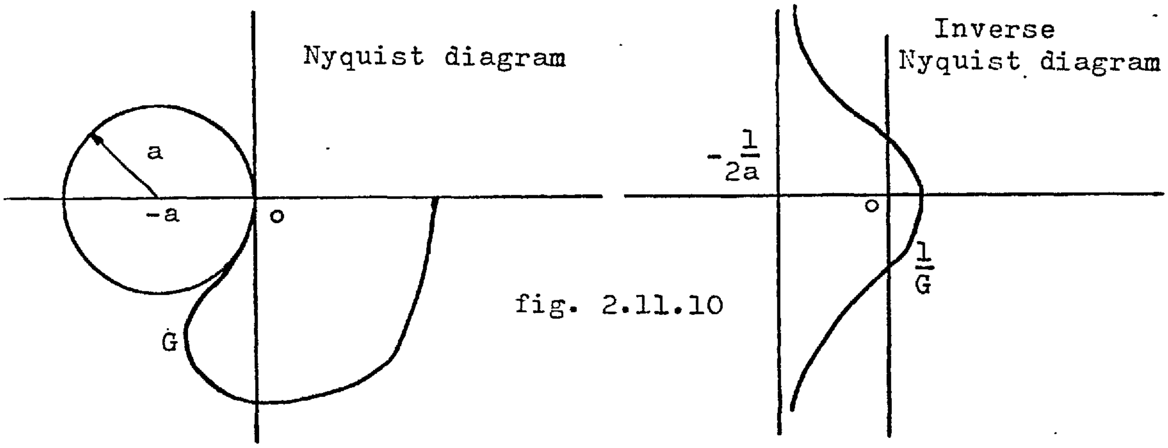


fig. 2.11.10

Hence an optimal $\frac{1}{a}$ on the inverse Nyquist diagram lies entirely to the right of the vertical line with abscissa $-\frac{1}{2}$. If $G \longrightarrow 0$ on the Nyquist diagram with asymptotic radius of curvature a, then $\frac{1}{G}$ is asymptotic to the vertical line $x = -\frac{1}{2a}$.

CHAPTER 3

RANDOM PROCESSES

3.1 Optimal Control of Stochastic Systems

While almost all of the results in this short chapter are known, it is included for two main reasons. One is to show how abstract concepts can deal with stochastic processes quite concisely. The other reason for this chapter is to develop the problem of identification, which provides a motivation for Chapter 4.

Loeve [L 4; Ch. 8] develops conditions under which stochastic processes (not necessarily stationary, Gaussian, or with zero mean) can be considered as elements of Hilbert spaces. The inner product can be defined by an expectation operation (with perhaps time integration). For families of stochastic processes which constitute a Hilbert space, the development of Section 2.2 is valid. In particular, for the stochastic control problem, the equations

$$J = \langle y, Q\, y \rangle + \langle u, R\, u \rangle \qquad 3.1.1$$

$$y = y_o + W\, u \qquad 3.1.2$$

$$\pi(R\, u + W^* Q\, y) = 0 \qquad 3.1.3$$

where y_o, y and u are interpreted as stochastic processes, remain meaningful and valid. The π of equation 3.1.3 is included for greater generality, and represents a projection operation of the gradient into the space of allowable control variation. For deterministic problems, π projects the gradient onto the subspace of time functions which are identically zero for negative time. For stochastic problems, since usually only the past and present data (noisy measurements) are given, the projection operator projects the gradient into the space spanned by these measurements. The projection π enables the notion of causality to be used, as developed in Section 1.8.

The operator R is taken to be instantaneous (and hence causal) with bounded inverse R^{-1}. Then from 3.1.3

$$\begin{aligned} u &= - R^{-1} \pi (W^{*} Q y) \\ &= - \pi (R^{-1} W^{*} Q y) \ , \qquad 3.1.4 \end{aligned}$$

where $u \in \mathcal{R}(\pi)$ by assumption. In general, the only variable available for control synthesis is the past history of a measurement vector m, which consists of a linear operation on the outputs, corrupted with additive independent measurement noise v. i.e.

$$m = C y + v \quad . \qquad 3.1.5$$

Hence, from 3.1.4, u is the optimal <u>linear</u> estimate of $- R^{-1} W^{*} Q y$, even if y_o, and hence u and y, are non-Gaussian processes. To specify the projection explicitly, we write 3.1.4 as

$$u = -\pi_{|m} \left(R^{-1} W^{*} Q y\right) .$$

However, this projection can be arrived at in two steps.

$$u = -\pi_{|m} \left(\pi_{\substack{|m \\ y}} \left(R^{-1} W^{*} Q y\right)\right) . \qquad 3.1.6.$$

Let
$$z = \pi_{\substack{|m \\ y}} \left(R^{-1} W^{*} Q y\right)$$

$$= \pi_{|y} \left(R^{-1} W^{*} Q y\right)$$

$$= H_1 y \qquad 3.1.7$$

where H_1 is a causal operation on y. But

$$u = -\pi_{|m} \left(H_1 y\right)$$

$$= -H_1 \pi_{|m} (y) \qquad 3.1.8$$

due to the commutation of projection and causal operators;

i.e.
$$u = -H_1 \hat{y}$$

$$= -H_1 H_2 m , \qquad 3.1.9$$

where
$$\hat{y} = H_2 m \qquad 3.1.10$$

is the optimal linear estimate of y from the data m, and H_2 is causal. This derivation shows the application of the certainty equivalence principle, that if variables are not measured directly, they should be replaced by their optimal linear estimates.

For Gaussian processes, optimal linear estimation and optimal estimation are equivalent; i.e. projectioning and conditioning are equivalent for Gaussian processes [L 4; p.462]. Hence

$$R\,u + \underset{|m}{E}\, W^{*} Q\, y = 0 \,. \qquad 3.1.11$$

Also

$$u = \underset{|m}{E}\, (R^{-1} W^{*} Q\, y)$$

$$= \underset{|m}{E}\, (\underset{|y}{E}\, R^{-1} W^{*} Q\, y) \,. \qquad 3.1.12$$

If y represents all of the states x of a continuous system represented in a finite state-space form, then

$$u = \underset{|m}{E}\, (\underset{|x}{E}\, R^{-1} B^{T} P\, x)$$

$$= \underset{|m}{E}\, (R^{-1} B^{T} P\, x)$$

$$= R^{-1} B^{T} P\, \hat{x} \qquad 3.1.13$$

where P is calculated from the Riccati equation 2.6.6 independently of the stochastic processes involved, and $\hat{x}$ is an estimate of x independent of the control problem. Hence the problem reduces to finding H_1 and H_2, or $H_1 H_2 = H$ in equations 3.1.9-10. To compute these operations, knowledge of statistical parameters are required. In particular, for Gaussian processes, knowledge of means, and auto- and cross-covariances completely specify the processes.

3.2 Correlation

Auto- and cross-correlation functions, and their transforms (if defined) have proved very useful in the study of linear systems. For deterministic signals, the cross-correlation of two (vector) time functions can be defined as an integral:

$$r_{yu}(t) = \int_{-\infty}^{\infty} y(\tau)\, u^T(\tau - t)\, d\tau \quad . \qquad 3.2.1$$

For stationary zero-mean stochastic processes, the expectation operator (which is just an integral with respect to probability measure) is used:

$$r_{yu}(t) = E\, y(\tau)\, u^T(\tau - t) \quad . \qquad 3.2.2$$

It is preferable to have an abstract definition of correlation, applicable to elements of a general Hilbert space, and this can be obtained by the use of the development in Section 1.2. If W is a map from $\mathcal{H}_u$ to $\mathcal{H}_y$, where if $u \in \mathcal{H}_u$, $y \in \mathcal{H}_y$

$$y = W u$$

then, if also $z \in \mathcal{H}_y$, we can form the inner-product

$$\begin{aligned} f &= \langle z, y \rangle \\ &= \langle z, W u \rangle \quad . \end{aligned}$$

However, if u is interpreted as a mapping from the Banach algebra $\mathcal{J}$ into $\mathcal{H}_y$, then we can define its conjugate mapping from $\mathcal{H}_y$ into $\mathcal{J}^*$, the space of continuous functionals on $\mathcal{J}$.

i.e.

$$f = \langle z, W u \rangle \tag{3.2.3}$$

$$= \langle z u^*, W \rangle \quad . \tag{3.2.4}$$

The functional $z u^* = r_{zu}$ is defined as the cross-correlation of z and u.

<u>Theorem 3.2.1</u>

$$r_{yu} = (r_{uy})^* \tag{3.2.5}$$

<u>Proof</u>: r_{yu} is shown to represent a linear operation, and the * operation denotes its adjoint. Then consider, $W \in \mathcal{J}_{uy}$ and $H \in \mathcal{J}_{yy}$. Then the inner product

$$f = \langle H y, \; W u \rangle \tag{3.2.6}$$

on $\mathcal{H}_y$ is defined.

But

$$f = \langle (H y) u^*, W \rangle$$

$$= \langle H(y u^*), W \rangle \quad . \tag{3.2.7}$$

But $y\,u^*$ can be considered a linear operator on H. Hence, by the rule for definition of conjugates

$$f = < H, W(y\,u^*)^* > . \qquad 3.2.8$$

However, from 3.2.6

$$f = < H, W\,u\,y^* > . \qquad 3.2.9$$

Since W and H are arbitrary, we must have

$$u\,y^* = (y\,u^*)^* .$$

i.e. $$r_{uy} = (r_{yu})^* .$$

Returning to the control problem, the correlations of 3.1.2 with respect to the variables y_o, y and u yield

$$r_{yy_o} = r_{y_oy_o} + W\,r_{uy_o} \qquad 3.2.10$$

$$r_{yy} = r_{y_oy} + W\,r_{uy} \qquad 3.2.11$$

$$r_{yu} = r_{y_ou} + W\,r_{uu} . \qquad 3.2.12$$

Since $u \in \mathcal{R}(\pi)$, and hence $H\,u \in \mathcal{R}(\pi)$ when H is causal, it is orthogonal (independent for Gaussian processes) to the error in projection. So

$$
\begin{aligned}
< R\,u + \pi(W^{*}\,Q\,y),\ H\,u > \; &= \; < R\,u + W^{*}\,Q\,y,\ H\,u > \\
&\quad + < \pi(W^{*}\,Q\,y) - W^{*}\,Q\,y,\ H\,u > \\
&= \; < R\,u + W^{*}\,Q\,y,\ H\,u > \; + \; 0 \\
&= \; < R\,u\,u^{*} + W^{*}\,Q\,y\,u^{*},\ H > .
\end{aligned}
$$

Since H is arbitrary

$$R\,r_{uu} + W^{*}\,Q\,r_{yu} = 0 \quad . \qquad 3.2.13$$

This analysis applies whether stochastic or deterministic problems are being considered, and shows that the solution of the problem depends purely on the correlation functions, for equations 3.2.10-13, together with

$$u = -H\,m \qquad 3.2.14$$

completely specify the solution. Hence if deterministic and stochastic problems have the same correlations, the filter H will be the same, and so solution of the stochastic problem is reduced to finding an equivalent deterministic problem.

For systems conveniently describable by the state-space approach, the separation of control and estimation provides a very useful method of systems design. For stationary, zero-mean systems only empirically known in terms of weighting functions and measured correlations the following method is proposed. Consider Figure 3.2.1.

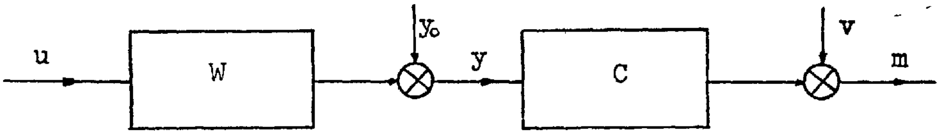


fig. 3.2.1

For convenience, we will take C = I. This is not necessary, however, and C can be included throughout if desired.

Then
$$m = y_o + v + W u$$
$$= d + W u \quad . \qquad 3.2.15$$

However, we assume that

$$r_{y_o v} = 0 \qquad 3.2.16$$

and
$$r_{vv} = \sigma_{vv} \, \delta \qquad 3.2.17$$

where σ_{vv} is assumed a P.D. covariance matrix, and δ represents the unit operator (dirac delta). (This is not strictly a functional in $\mathfrak{J}^*$, but belongs to the wider class $\mathfrak{J}^f$.) But from 3.1.3

$$R u + \pi_{|m} (W^* Q y)$$

$$= R u + \pi_{|m} W^* Q(m - v) \quad .$$

But since v is white, and W^* is an anticipatory operator

$$\pi_m \, W^* Q \, v = 0 \quad . \qquad 3.2.18$$

i.e.

$$R \, u + \pi_m \, W^* Q \, m = 0 \, . \qquad 3.2.19 \, .$$

Hence 3.2.15 and 3.2.19 define a dynamic equation and a gradient equation. The method then reduces to finding a deterministic matrix signal D with auto-correlation r_{dd}, and performing an optimal control calculation as described in Chapter 2. With the appropriate trajectories, a filter H is calculated via the algorithms of Section 2.10. Note that if v is white, the matrix D will have an impulse in it, but since this cannot be corrected by any control, it will not affect the solution. In fact, D can always be chosen minimum phase, as shown in the next section, thus eliminating some of the difficulties associated with the stability of H.

If

$$\left.\begin{aligned} (I + L) r \, (I + L^*) &= r_{y_o y_o} + r_{vv} \\ &= r_{dd} \, , \end{aligned}\right\} \quad 3.2.20$$

where $(I+L)^{-1}$ is stable, then L is a suitable matrix signal to use in the algorithms of 2.10. However, the I of 3.2.20 simplifies calculation, for we will wish to find H such that

$$U = H(I + Y)$$

$$= H + H Y .$$

i.e. $$H = U - H Y \qquad 3.2.31$$

and this equation can be used recursively to calculate H.

3.3 Optimal Filtering and Duality

In the last two sections, it was shown that for Gaussian processes, optimal filtering can be separated from optimal control, leading to a certainty equivalence principle. Filtering also has important applications in its own right. However, when the filtering problem is formulated in abstract notation, a duality with the control problem becomes immediately obvious. We shall consider the least squares filtering problem; i.e. we wish to find the minimum variance estimate of a noisy signal. Consider the system of Figure 3.3.1.

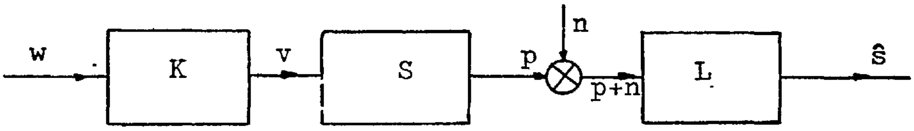


fig. 3.3.1

$$r_{ww} = I \qquad 3.3.1$$

$$r_{nn} = \sigma_{nn} I \qquad 3.3.2$$

$$r_{nw} = 0 . \qquad 3.3.3$$

If $s = S\,w$, where S is a causal operator & K is an instantaneous operator, we wish to choose a causal filter L such that

$$\hat{s} = L(p + n)$$

$$= L(S\,K\,w + n)\ , \qquad 3.3.4$$

where $$J = \langle s - \hat{s},\ s - \hat{s} \rangle \qquad 3.3.5$$

is to be minimised with respect to L.

But $$J = \langle s - L(p + n),\ s - L(p + n) \rangle$$

$$= \langle s - Lp,\ s - Lp \rangle - 2\langle s - Lp,\ Ln \rangle + \langle Ln,\ Ln \rangle .$$

However, since n is uncorrelated with w, and hence with s or p,

$$\langle s - Lp,\ Ln \rangle = 0 \qquad 3.3.6$$

i.e. $$J = \langle s - Lp,\ s - Lp \rangle + \langle Ln,\ Ln \rangle\ . \qquad 3.3.7$$

Let $$YKw = s - Lp$$

$$= Sw - LSKw\ .$$

i.e. $$YK = S - LSK\ . \qquad 3.3.8$$

Then $$J = \langle YKw,\ YKw \rangle + \langle Ln,\ Ln \rangle$$

$$= \langle YKww^{*},\ YK \rangle + \langle Lnn^{*},\ L \rangle$$

$$= \langle YK,\ YK \rangle + \langle L\sigma_{nn},\ L \rangle$$

$$= < K^* Y^*, K^* Y^* > + < L^*, \sigma_{nn} L^* > \qquad 3.3.9$$

$$= < Y^*, \sigma_{vv} Y^* > + < L^*, \sigma_{nn} L^* > \qquad 3.3.10$$

where σ_{vv} and σ_{nn} are self-adjoint.

Also $$K^* Y^* = S^* - K^* S L^* \qquad 3.3.11$$

where $Y_o = S K^{-1}$.

Equations 3.3.10, 3.3.11 represent (in terms of operators) the equations for an optimal control problem. However, the operators are represented in terms of their adjoints (duals). That is L^* takes the place of U and S^* the place of W. By direct analogy with the control problem, the filtering problem reduces to the spectral factorisation of the operator

$$A = \sigma_{nn} + S \sigma_{vv} S^* . \qquad 3.3.12$$

i.e. $$(I + L)\sigma \ (I + L^*) = \sigma_{nn} + S \sigma_{vv} S^* . \qquad 3.3.13$$

For multivariable systems in the time (frequency) domain, duality implies transposition of matrices and reversal of the sign of time (frequency).

For Markov processes, Kalman and Bucy have shown that the filtering may be performed recursively [KB 1]. In particular, for the noise model of Figure 3.3.2, the filter is shown diagramatically.

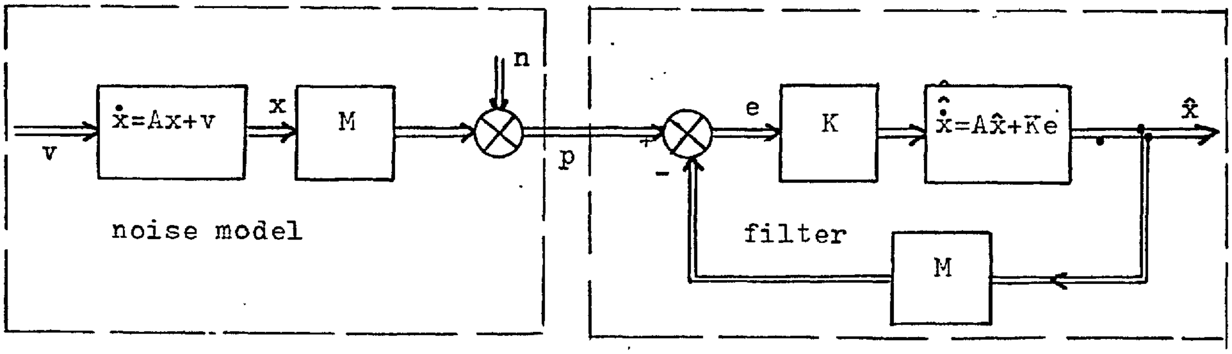


fig. 3.3.2

$$E\, v(t)\, v^T(t-\tau) = \sigma_{vv}\, \delta(t-\tau)$$

$$E\, n(t)\, n^T(t-\tau) = \sigma_{nn}\, \delta(t-\tau) \qquad 3.3.14$$

$$E\, v(t)\, n^T(t-\tau) = 0 \quad .$$

$$K(t) = S(t)\, M^T \sigma_{nn}^{-1} \qquad 3.3.15$$

where $S(t)$ is the optimal covariance matrix of the estimation error, given by

$$\frac{dS}{dt} = A\,S + S\,A^T - S\,M^T \sigma_{nn}^{-1} M\,S + \sigma_{vv} , \qquad 3.3.16$$

where the initial $S(0)$ is given. This equation is a matrix Riccati equation in forward time, dual to the backward time equation arising in the control problem. The two equations have many dual properties. In particular, the filtering equation can also solve the spectral factorisation problem in a similar way to the results of Section 2.9.

Discrete systems have similar equations.

This filter forms an integral part of the control system for noisy stochastic systems. The overall controller-estimator diagram (in the stationary frequency-domain case) becomes the system in Figure 3.3.3.

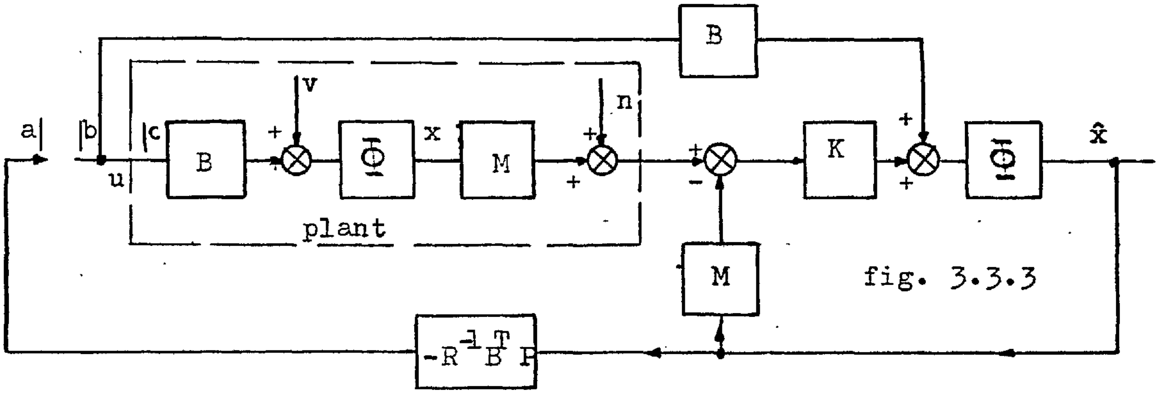


fig. 3.3.3

$$\Phi(s) = (sI - A)^{-1} .$$

Systems designed using the Kalman filter as a control compensator have been extensively investigated. In particular, the filter itself is always stable, as in the overall closed loop system. An important property is the following.

<u>Theorem 3.3.1</u>: The overall loop gain from b to a is the optimal gain G(s) obtained from considering the control problem alone, and this is independent of K and M.

<u>Proof</u>: $$G(s) = R^{-1} B^T P \Phi(s) B . \qquad 3.3.17$$

But the transfer T from b to a is given by

$$-T = R^{-1} B^T P H$$

where H is the transfer from u to $\hat{x}$. But, neglecting noise,

$$\hat{x} = \Phi K(M \Phi B u) + \Phi B u - \Phi K M \hat{x} .$$

$$(I + \Phi K M)\hat{x} = (K M \Phi B + \Phi B)u .$$

i.e. $$H = (I + \Phi K M)^{-1} (K M \Phi B + \Phi B)$$

$$= (I + \Phi K M)^{-1}(\Phi K M + I) \Phi B$$

$$= \Phi B .$$

i.e. $$-T = R^{-1} B^T P \Phi(s) B = G(s).$$

If the loop at a b is closed, but opened at b - c, then, with the plant removed, there seems to be no guarantee that the resulting compensator (which is the one that is actually built, and corresponds to the filter designed by the method above) is stable. That is, the system represented in Figure 3.3.4 may not be stable.

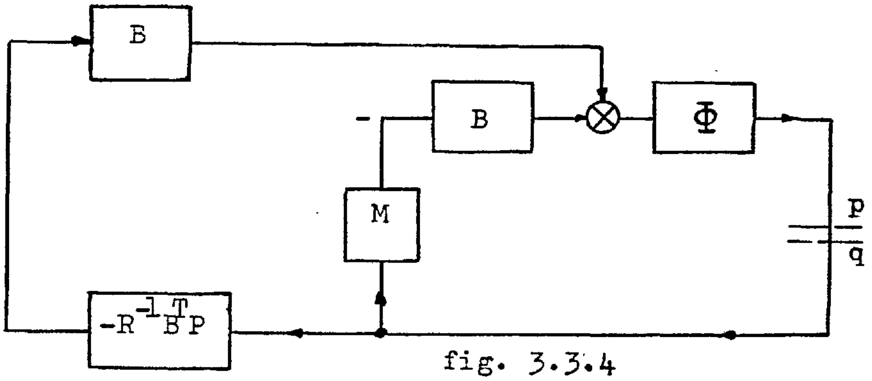


fig. 3.3.4

The loop gain from q to p is given by

$$L = \Phi (K M + B R^{-1} B^T P) \qquad 3.3.18$$

For stability, $(I + L)^{-1}$ should be stable .

Let $$S = I + L$$

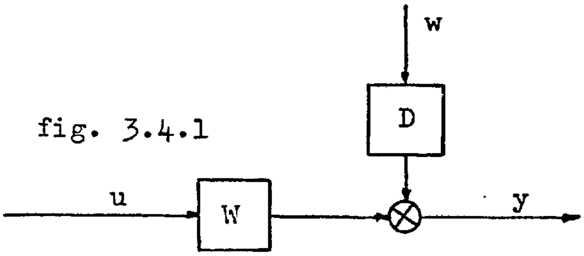
$$= I + \Phi K M + \Phi B R^{-1} B^T P . \qquad 3.3.19$$

Then S^{-1} is not guaranteed stable.

3.4 Identification and Estimation

The development of the last few sections has shown that some stochastic problems can be reduced to the solution of deterministic ones. In fact, considering Figure 3.4.1,

fig. 3.4.1



knowledge of the operators W and D determine the solution of the control problem. Correlation measurements on an actual system enable

W, and D (to within an all-pass system), to be determined. Of course, in any actual statistical experiment, W and D can only be determined to a finite error variance, due to the essential finiteness of the test. It is generally assumed that the input u is accessible, but that the input w is inaccessible, and this makes W essentially easier to estimate than D.

A popular method of estimating W is to let u be white noise of unit variance, and measure the cross correlation of the output y with u. Let

$$u\,u^* = I \quad . \qquad 3.4.1$$

$$u\,w^* = 0 \quad . \qquad 3.4.2$$

Then we can state

Theorem 3.4.1: $W = y\,u^* \quad .$ 3.4.3

Proof: $y = d + W\,u$ 3.4.4

where $d = D\,w \quad .$ 3.4.5

Let H be an operator which maps white noise into a valid element of a Hilbert space. W also has this property. Then, formally,

$$<y, H\,u> = <y\,u^*, H> \quad .$$

But $<y, H\,u> = <d + W\,u, H\,u>$

$$= < d\, u^* + W\, u\, u^*,\ H > .$$

$$= 0 + < W\, I,\, H >$$

$$= < W,\, H > .$$

Since H is arbitrary,

$$W = y\, u^*.$$

Hence the problem of estimating W reduces to estimating r_{yu}, the cross correlation of y and u. Since u is accessible, variance reduction Monte-Carlo techniques can be used. In any practical control system, there will be inherent non-linearities in the plant, and the estimate of W will depend on the actual variance of u chosen, as well as its distribution (not necessarily Gaussian). This estimation technique then may provide a convenient method of statistical linearisation.

Since w is inaccessible, the estimation of D is a more difficult problem. It is then only possible to estimate D to within an all-pass filter, and the natural choice to make is that D should be minimum phase. For $u = 0$,

$$r_{yy} = r_{dd}$$

$$= D\, w \,.\, w^* D^*$$

$$= D\, D^* \qquad 3.4.6$$

Hence a spectral factorisation of r_{dd} is needed to determine the operator D, whose inverse is assumed stable. Since, in general, d has a white noise component, r_{dd} will be bounded below, and so D will have a bounded inverse. Hence, any of the methods of Section 2.9 are immediately applicable.

In general, we can only find an estimate of r_{yy} due to a finite set of measurements. The following development, for the particular case of a stationary discrete time system, shows that in the case of a finite (though large) data set, the spectral factorisation method produces a good estimate. The technique of maximum likelihood estimation is applied to the data set to determine an estimate of the weighting matrix sequence D_j. That is, we let

$$F = D^{-1} \qquad 3.4.7$$

and construct F such that

$$w = F y . \qquad 3.4.8$$

Consider the system in Figure 3.4.2.

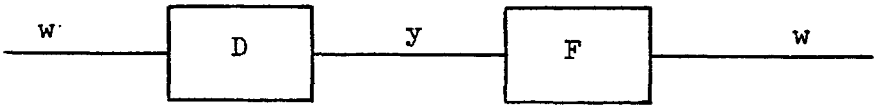


fig. 3.4.2

$$r_{yy_j} = \sum_{i=0}^{\infty} D_i D_{i-j}^T . \qquad \forall\ j,\ \text{integral.} \qquad 3.4.9$$

$$y_k = \sum_{j=-\infty}^{k} D_{k-j} w_j \qquad 3.4.10$$

$$w_j = \sum_{i=-\infty}^{j} F_{j-i} y_i \qquad 3.4.11$$

$$\text{where} \quad \sum_{j=1}^{k} F_{j-i} D_i = \sum_{j=1}^{k} D_{j-i} F_i = I \delta_k \qquad 3.4.12$$

and δ_k is the Kronecker delta function.

Consider a sequence of N independent random vector variables w_r which depend on the parameters $(F_{ij})_k$ which we include in a vector θ. The sequence of vectors w_r can also form a partitioned vector w. Then the probability of w, given θ can be written

$$p(w|\theta) = \prod_{r=1}^{N} p(w_r|\theta) \qquad 3.4.13$$

since the w_r (being white) are independent. Now define

$$L(w|\theta) = - \ln p(w|\theta)$$

$$= - \sum_{r=1}^{N} \ln p(w_r|\theta) . \qquad 3.4.14$$

By the maximum likelihood technique, $p(w|\theta)$ is maximised with respect to θ. This is equivalent to minimising L, the likelihood function, and the value of θ that produces this minimum is called the maximum likelihood estimator (M.L.E.). However, we can use the following:

Theorem 3.4.2: If N independent observations w_i, $i = 1 \ldots N$, with $p(w_i \theta)$ known, are available, then provided the prior density $p(\theta)$ is nowhere zero, the posterior density of θ for large N is approximately normal with

$$\text{mean} \quad \hat{\theta} = \left\{ \theta : \frac{\partial L}{\partial \theta} = 0 \right\} \tag{3.4.15}$$

$$\text{and variance } \sigma_{\theta\theta} = L^{-1}_{\theta\theta}(w|\hat{\theta}) \ . \qquad [\text{M } 2]. \tag{3.4.16}$$

If w_i is assumed Gaussianly distributed, with zero mean, and unit matrix covariance, then

$$\prod_{r=1}^{N} p(w_r|\theta) = \prod_{r=1}^{N} \frac{1}{(2\pi)^{p/2}} \exp[-\tfrac{1}{2} w_r^T w_r]$$

$$= \frac{1}{(2\pi)^{Np/2}} \exp[-\tfrac{1}{2} \sum_{r=1}^{N} w_r^T w_r] \ . \tag{3.4.17}$$

$$\text{i.e.} \qquad L(w|\theta) = \frac{N \, p \ln 2}{2} + \tfrac{1}{2} \sum_{r=1}^{N} w_r^T w_r \ . \tag{3.4.18}$$

For $\hat{\theta}$ to be the M.L.E.,

$$\frac{\partial L}{\partial \theta|\hat{\theta}} = 0 \tag{3.4.19}$$

i.e. each matrix

$$\frac{\partial L}{\partial F_k} = 0 \ , \quad \text{for each } k = 0, 1, 2, \ldots \qquad 3.4.20$$

But
$$\frac{\partial L}{\partial F_k} = \sum_{r=1}^{N} \frac{\partial L}{\partial (w_r^T w_r)} \frac{\partial (w_r^T w_r)}{\partial F_k}$$

$$= \frac{1}{2} \sum_{r=1}^{N} \frac{\partial (w_r^T w_r)}{\partial F_k} \ . \qquad 3.4.21$$

However, from 3.4.11

$$w_r = \sum_{k=-\infty}^{r} F_{r-k} y_k$$

$$= \sum_{k=o}^{\infty} F_k y_{r-k} \ . \qquad 3.4.22$$

$$\therefore \quad w_r^T w_r = (\sum_{i=o}^{\infty} y_{r-i}^T F_i^T)(\sum_{k=o}^{\infty} F_k y_{r-k})$$

$$= \sum_{i=o}^{\infty} \sum_{k=o}^{\infty} (y_{r-i}^T F_i^T F_k y_{r-k})$$

$$= \text{tr.} \sum_{i=o}^{\infty} \sum_{k=o}^{\infty} (F_i y_{r-i} y_{r-k}^T)^T F_k . \qquad 3.4.23$$

Hence
$$\frac{\partial (w_r^T w_r)}{\partial F_k} = 2 \sum_{i=o}^{\infty} F_i y_{r-i} y_{r-k}^T \ . \qquad 3.4.24$$

Hence if $$\frac{\partial L}{\partial F_k} = 0 \quad \forall \; k = 0, 1, 2, \ldots$$

$$\sum_{r=1}^{N} \sum_{i=o}^{\infty} F_i \, y_{r-i} \, y_{r-k}^T = 0 \quad . \qquad 3.4.25$$

Changing the order of summation

$$\sum_{i=o}^{\infty} F_i \sum_{r=1}^{N} y_{r-j} \, y_{r-k}^T = 0 \qquad \text{for } k = 0, 1, 2, \ldots \qquad 3.4.26$$

If the stationary process y is assumed ergodic, then for large N,

$$\frac{1}{N} \sum_{r=1}^{N} y_{r-i} \, y_{r-k}^T = \hat{r}_{yy_{k-i}} \qquad 3.4.27$$

provides an estimate of the auto-correlation matrix sequence of y. Hence the asymptotic result (equation 3.4.6) is seen to hold for an infinite data set.

i.e. $$D^{-1} r_{yy} = D^{*} = 0 \qquad \text{for positive time.}$$

However, for a finite data set, the analysis provides an estimate of the error variances, though the analysis becomes complicated. The set of equations

$$\sum_{i=o}^{\infty} F_i \, \hat{r}_{k-i} = 0 \qquad k = 0, 1, 2 \ldots \qquad 3.4.28$$

is an infinite set in an infinite of unknowns. However, there exists an indeterminacy (apart from an arbitrary unitary transformation) that is resolved by the fact that

$$r_o = \sum_{i=o}^{\infty} D_i D_i^T . \qquad 3.4.29$$

The following method of solution is proposed. First scale the F sequence by F_o^{-1}; i.e.

$$\begin{aligned} E_o &= F_o^{-1} F_o = I \\ E_1 &= F_o^{-1} F_1 \qquad\qquad 3.4.30 \\ E_2 &= F_o^{-1} F_2 \qquad \ldots \end{aligned}$$

With $E_o = I$, 3.4.28 is only valid for $k = 1, 2, 3 \ldots$ In full, the following set of equations is obtained:

$$\left.\begin{aligned} E_1 r_o + E_2 r_1^T + E_2 r_2^T + \ldots &= - r_1 \\ E_1 r_1 + E_2 r_o^T + E_3 r_1^T + \ldots &= - r_2 \\ E_1 r_2 + E_2 r_1 + E_3 r_o + \ldots &= - r_3 \end{aligned}\right\} 3.4.31$$

i.e. $$\left(E_1 \mid E_2 \mid E_3 \mid \ldots\right) \left(\begin{array}{c|c|c|c} r_o & r_1^T & r_2^T & \ldots \\ \hline r_1 & r_o & r_1^T & \ldots \\ \hline r_2 & r_1 & r_o & \ldots \\ \hline \vdots & & & \vdots \end{array}\right) = - \left(\begin{array}{c} r_1 \\ \hline r_2 \\ \hline r_3 \\ \hline \vdots \end{array}\right) \qquad 3.4.32$$

This set of equations can be solved by the Cholesky factorisation technique of Section 2.9. For the single variable case, a similar algorithm has been proposed by Levinson[W 2], which approximates the solution by taking larger and larger top corner blocks of equation 3.4.32. This was programmed and proved to be excellent.

Having found the sequence $I, E_1, E_2 \ldots$, this is inverted by back substitution to provide a solution sequence $I, C_1, C_2 \ldots$,

where $$C_i = -(C_{i-1} E_1 + C_{i-2} E_2 + \ldots + E_i) . \qquad 3.4.33$$

By virtue of Theorem 2.10.1, this C sequence can be used directly to determine the feedback compensator, by solving an optimal control problem using one column of C at a time.

To determine the scaled sequence D_i (which can only be unique to within an arbitrary unitary matrix), the following method of solution is proposed. Let

$$M = I + \sum_{i=1}^{\infty} C_i C_i^T . \qquad 3.4.34$$

Then, from equation 3.4.29

$$\hat{r}_o = D_o M D_o^T . \qquad 3.4.35$$

If both M and r_o (both symmetric) are factorised via either Cholesky triangulation, or the square root algorithm, such that

$$\hat{r}_o = B\,B^T$$ 3.4.36

and $$M = A\,A^T$$

then, equating factors,

$$B = D_o\,A\ .$$

$$\therefore \qquad D_o = B\,A^{-1}$$ 3.4.37

Hence the C sequence can be rescaled to produce

$$D_i = D_o\,C_i\ .$$ 3.4.38

CHAPTER 4

APPROXIMATION, SENSITIVITY AND BOUNDS

4.1 Design Practicalities

The previous chapters have considered in detail the properties of optimal control, and closed loop design based on optimal control. For control system design, the theory of optimal control has two main uses:

1. To show exactly what can be achieved for a fixed performance index, and provide a lower bound for comparison with a sub-optimal design.

2. As a design method in itself.

In this chapter, the calculation of optimal control laws as a design method will be adopted.

Inherent in engineering design criteria, though often not explicitly specified in quantitative terms, is that the resultant control system should be as simple as possible. The designer is often prepared to allow some degradation in the explicit performance index to achieve this. Another important, but often unstated, criterion is that the control law should not be too sensitive to the actual plant it controls.

For linear systems, a simple design usually implies either gain feedback from a few important outputs, or a combination of these with compensators consisting only of a small number of integrators and gains. Pure delays may or may not be tolerated as compensator elements, depending on whether the realisation of the system is by analog or digital means. However, one only obtains simple designs from optimal control theory, when

(a) a quadratic cost criterion with constant weighting matrices is used, over an infinite time interval, and

(b) the dynamic system is linear, time-invariant, and has a state-vector representation, where the dimension of the state is small.

In the above case, either pure gain feedback from all the states can be designed, if all the states are available for measurement, or a Kalman filter type compensator can be used to estimate those states that are unavailable. This type of design has been well investigated in the literature. However, when the state-space representation is not of small dimension, then simple designs no longer result from the theory, and designs of the type described in Chapters 2 and 3, where dynamic filters are synthesised by numerical convolution of impulse responses with measured signals, become more attractive to implement.

Classical feedback theory, however, shows that good designs can often be achieved by simple control structures, even for complex plants.

This results from the fact that usually only a few states are "dominant" and the remainder contribute very little to the performance. However, to achieve the property of infinite gain margin, as predicted from optimal control (Section 2.11), all states must be fed back. Since there is no physical system whose gain margin is actually infinite, true optimal control can never really be implemented. This implies that under high values of gain, simple models may cease to be valid.

In the light of Section 3.4, the exact transfer characteristics of a physical plant can never be measured exactly. It seems reasonable to try and approximate physical plants by low-dimensional state-space models, though the mathematical justification for this constitutes a field for further research. We shall sidestep the investigation of the errors involved in the approximation of a physical plant by a simple model, and confine the discussion to the approximation of a complex linear model by a simple one.

What we wish to do in this chapter is the following. We assume that a complex model is given, and that an approximation is made according to some engineering criterion to obtain a simple model. Algorithms for making approximations have been developed for many years, and are still a research topic. Some of these are briefly described in the next section. Now if an optimal control problem is solved using this simple model to obtain a set of feedback gains, or a compensator design, we try to answer the following questions.

1. Is this design good when applied to the complex model?

2. How do we choose the matrices Q and R such that good designs will result?

Truxal [T2] gives a good resumé of the difficulties of estimation, and the associated control design problem. The reason for the design difficulty is that while open loop characteristics of two systems may be approximately alike, the characteristics under closed loop control may be entirely different. In fact, one system may have excellent properties while the other is unstable. From an engineering viewpoint, this is simply explained. At high frequencies, there is usually more phase shift in the actual plant than in the model. This is unimportant in the open loop system, since it does not contribute very much to the transfer characteristics when the gain is low, and hence is difficult to estimate. However, when large gain is applied, the actual gain at -180° phase shift can become greater than one, causing instability. Hence, we are in the unfortunate position of having the control design depend on the system model, and the system model depend on the control design. The more we want the system to do, the better our model should be, such that the open-loop approximations remain valid in the closed loop system. The following example can illustrate these points.

<u>Example 4.1.1</u>: Complex model: $\left(\frac{10 - s}{10 + s}\right) \cdot \frac{1}{s}$

Simple model: $\frac{1}{s}$

Comparison of the transient responses of the two systems (Figure 4.1.1) shows that, on a large time scale, there is little difference between the systems.

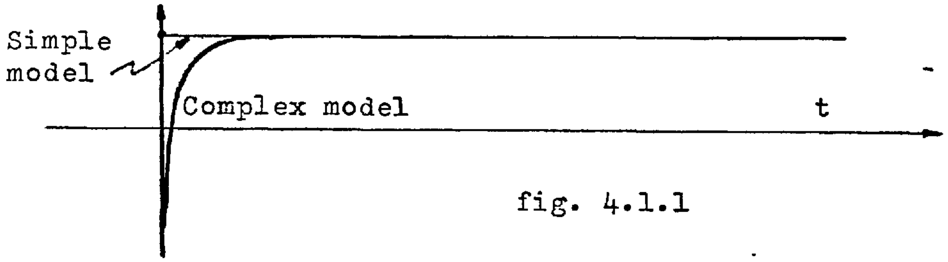


fig. 4.1.1

A gain of 10 causes instability in the complex plant while the simple plant has a good quick response. However, small gains from 0 - 3 show similar closed loop responses from both systems.

Having performed a control design on a simple model, it is important to be able to estimate the degradation in performance of the more complex system. Since gain that is too high can cause instability, it is useful to have an estimate of an upper bound on gain in terms of modelling errors. If an optimal control calculation is performed, limits on the Q and R matrices are preferable.

Throughout the remainder of this thesis, only time-invariant systems will be considered, although those results presented in abstract terms can apply to the time-varying case. In particular, continuous time (multivariable) systems are taken as the prime examples.

Also in the following discussion, the philosophy behind the weighting function approach will become apparent. For the state-space description, small changes in modelling only result from small parameter changes in the system matrices, whereas small delays, extra phase shift, and especially corners in transient response produce large structural changes in the model, since the dimension of the state space must be increased. However, most small system perturbations of this kind reduce only to small changes in the weighting function, which proves more versatile in this context.

4.2 Approximation

The classical methods of approximation of systems were originally derived in the context of circuit synthesis. These methods, which usually work in terms of frequency response, attempt to fit either experimental frequency curves, or irrational frequency functions to simple rational models, which are transforms of state-space models. Horovitz [H 5; Ch. 12] discusses some of these methods from the point of view of the control engineer.

One very popular method is Padé approximation. Given a rational Laplace transform model, of a size determined by engineering judgement, the parameters are fitted to idealised frequency response data by expanding both models into a power series in s, and truncating after all the unknown coefficients have been uniquely evaluated. In the time domain, this technique is equivalent to fitting moments, and in this way is often used for fitting experimental transient responses. One advantage of this method is that only simple equations need be solved. However, sometimes this method does not give very good approximations at all, as Horowitz points out.

Instead of fitting slopes, curvatures, etc., at the frequency origin, other methods fit curves at isolated frequencies. Similar methods are also used in the time domain. Modern methods of approximation specify an error criterion, and use computational hill-climbers to minimise this, in terms of the model parameters. Minimum sum of square errors, minimum of maximum error, etc., are all popular criteria, and have their advantages and disadvantages. In a statistical framework, the maximum likelihood technique has been successfully applied by Astrom[A 2]to fit a rational discrete time model to noisy data without first solving the set of linear equations in the filter impulse sequence derived in Section 3.4.

Horowitz makes the observation that it is far more important to analyse the errors of approximation when the system is to be incorporated into a closed loop, than when used in an open loop fashion. This is the prime difference between control design and filter design. We have described an example in Section 4.1, where instability results when modelling errors are not taken into account. The following example shows that approximation may predict instability, when in fact there is none.

Example 4.2.1: Ref: [C 2; p.170]

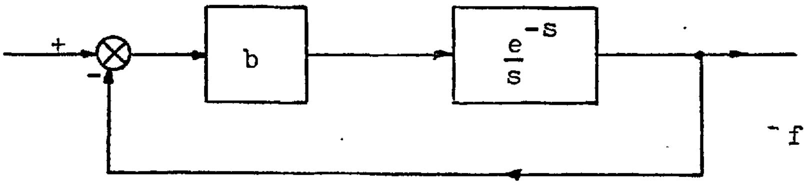


fig.4.2.1

It is known that this system is stable if the gain b lies in the region $0 < b < \frac{\pi}{2}$. However,

$$F(s) = \frac{\frac{b\ e^{-s}}{s}}{1 + \frac{b\ e^{-s}}{s}}$$

$$= \frac{b\ e^{-s}}{s + b\ e^{-s}} = \frac{N(s)}{D(s)} \quad .$$

$$D(s) = s + b\ e^{-s}$$

$$= b - (b - 1)s + \frac{b}{2}s^2 - \frac{b}{6}s^3 + \ldots$$

Truncation of this power series at any point indicates at least one pseudo-positive root - indicating instability.

A further point that arises in approximation theory is illustrated by a comment of Zadeh and Desoer [ZD 1; p.407]. They present an example of a system whose frequency domain curves differ by an arbitrarily small amount in amplitude, and yet whose time responses differ by an arbitrarily large amount at a particular point. This illustrates that the term "small" error should be interpreted in terms of the operator norm.

4.3 Sensitivity

Assume that a linear plant is given by equation 4.3.1.

$$y = y_o + W u \quad . \qquad 4.3.1$$

A control u is applied, which may or may not be implemented by a feedback law. A cost

$$J = \langle y, Q y \rangle + \langle u, R u \rangle \qquad 4.3.2$$

results from this control. Consider an arbitrary perturbation δW in the operator W, which may cause perturbations δu, δy, δy_o in u, y and y_o respectively. Other independent perturbations of these variables may also occur. Then the new cost becomes

$$\begin{aligned} J + \Delta J &= \langle y + \delta y, Q(y + \delta y) \rangle + \langle u + \delta u, R(u + \delta u) \rangle \\ &= \langle y, Q y \rangle + 2\langle y, Q \delta y \rangle + \langle \delta y, Q \delta y \rangle + \langle u, R u \rangle \\ &\quad + 2\langle u, R \delta u \rangle + \langle \delta u, R \delta u \rangle \ . \end{aligned}$$

i.e. $$\Delta J = 2\langle y,\ Q\ \delta y\rangle + 2\langle u,\ R\ \delta u\rangle + \langle \delta y,\ Q\ \delta y\rangle + \langle \delta u,\ R\ \delta u\rangle . \qquad 4.3.3$$

However $$\delta y = \delta y_o + \delta W\ u + W\ \delta u + \delta W\ \delta u\ . \qquad 4.3.4$$

$$\begin{aligned} \Delta J &= 2\langle y,\ Q\ \delta y_o\rangle + 2\langle y,\ Q\ \delta W\ u\rangle + 2\langle y,\ Q\ W\ \delta u\rangle \\ &\quad + 2\langle u,\ R\ \delta u\rangle + \langle \delta y,\ Q\ \delta y\rangle \\ &\quad + \langle \delta u,\ R\ \delta u\rangle + \langle y,\ Q\ \delta W\ \delta u\rangle \\ &= \delta J + \tfrac{1}{2}\,\delta^2 J\ . \qquad 4.3.5 \end{aligned}$$

$$\begin{aligned} \therefore \quad \tfrac{1}{2}\delta J &= \langle y,\ Q\ \delta y_o\rangle + \langle y,\ Q\ W\ \delta u\rangle + \langle u,\ R\ \delta u\rangle + \langle y,\ Q\ \delta W\ u\rangle \\ &= \langle y,\ Q\ \delta y_o\rangle + \langle W^{*}\ Q\ y,\ \delta u\rangle + \langle R\ u,\ \delta u\rangle + \langle Q\ y,\ \delta W\ u\rangle \\ &= \langle y,\ Q\ \delta y_o\rangle + \langle R\ u + W^{*}\ Q\ y,\ \delta u\rangle + \langle Q\ y\ u^{*},\ \delta W\rangle . \end{aligned}$$

4.3.6

<u>Theorem 4.3.1</u>: (Generalisation of a result by Pagurek [P2].)
The sensitivity of the cost to small perturbations of the system (both parameter and structural) is independent of whether feedback or open loop realisation is used, if the unperturbed system is optimal.

<u>Proof</u>: For optimal systems, the gradient

$$\begin{aligned} \hat{g} &= R\ \hat{u} + W^{*}\ Q\ \hat{y} \\ &= 0\ . \end{aligned}$$

Now δy_o does not depend on the control at all. If feedback is used, then δu depends on δW, and if not, then it is independent of δW. But since the sensitivity of cost with respect to δu is zero, by optimality, the sensitivity of cost to δW is independent of control variation, i.e.

$$\tfrac{1}{2}\,\delta J = \langle y, Q\,\delta y_o \rangle + \langle Q\,y\,u^*, \delta W \rangle . \qquad 4.3.7$$

In control design, the term of prime interest is the term due to δW alone, as δW affects the stability of the closed loop system. Using the same disturbance on different systems enables a good comparison of performance, and then $\delta y_o = 0$. In this case, the first order increase in cost for an optimal design is

$$\delta J = 2 \langle Q\,r_{yu}, \delta W \rangle \quad . \qquad 4.3.8$$

For continucus multivariable systems

$$\delta J = 2 \text{ tr} \int_0^\infty r_{yu}^T(\tau)\, Q\, \delta W(\tau)\, d\tau \quad . \qquad 4.3.9$$

The evaluation of δJ gives an estimate of the change in optimal cost due to a change in the weighting functicn. However, for the first order approximations to be applicable, we require

$$\delta^2 J \ll \delta J \;.$$

We shall make the rather drastic assumption that this is equivalent to

$$\delta J \ll \hat{J} \quad . \qquad 4.3.10$$

For simple systems, a reasonable engineering rule of thumb is

$$\frac{|\delta J|}{J} < 0.3 \qquad 4.3.11$$

for the first order estimate to be approximately valid. Using this rough criterion, the first order estimate of cost change can prove to be a useful check on an approximate design. If an optimal design is performed on a simplified model, to obtain a cost J, then the control law obtained may be implemented on a more realistic model, whence a change in cost ΔJ is obtained. If δJ is a good estimate of ΔJ, i.e. $\frac{|\delta J|}{J}$ remains small, then sensitivity provides a quick check on the applicability of the simple design. On the other hand, if $\frac{|\delta J|}{J}$ is large, then the actual closed loop system is invalidating the original approximations. A simple example illustrates these concepts.

Example 4.3.1: Complex model: $W(s) = \frac{1}{s(1 + 0.1s)}$

Simple model: $W(s) = \frac{1}{s}$

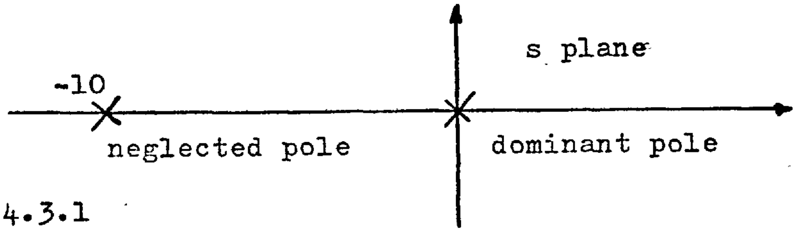


fig. 4.3.1

$$\delta W(s) = \frac{-0.1}{1 + 0.1s}$$

Consider the cost to a step disturbance; i.e.

$$y_o(s) = \frac{1}{s}$$

Let $$J = \int_{o}^{\infty} (y^2 + \frac{1}{k^2} u^2)\, dt \; .$$

Then, for the simple model, from the state-space analysis,

$$\hat{J} = \frac{1}{k}$$

and the optimal control is a negative feedback gain k. We now inquire into the validity of the application of this design to the complex plant. Using complex variable theory

$$\frac{1}{2}\delta J = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} y(-s)\, u(s)\, \delta W(s)\, ds$$

$$= \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{1}{-s + k} \cdot \frac{-k}{s + k} \cdot \frac{-0.1}{1 + 0.1s}\, ds \; .$$

From residue theory

$$\delta J = \frac{0.1}{1 + 0.1k} \quad .$$

$$\frac{|\delta J|}{J} = \frac{0.1k}{1 + o.1k} \quad .$$

For small k, $\frac{|\delta J|}{J}$ is small; the original dominant pole remains dominant, and the first order approximations remain valid. In this example, there is a large degree of stability margin, and values of k up to about 10 still produce acceptable responses from the second order plant, though the first order approximation is no longer valid.

The value $\frac{|\delta J|}{J} = 0.3$

is obtained for $k \sim 5$, the case of critical damping. Figure 4.3.2 plots $\frac{|\delta J|}{J}$ vs. k, and the actual $\frac{\Delta J}{J}$ vs. k. The method may also be used to check effects due to discarded cross-coupling terms in multivariable systems.

Example 4.3.2: Consider the system of Figure 4.3.3. This is simplified to the system

$$W(s) = \begin{pmatrix} \frac{1}{s} & 0 \\ 0 & \frac{1}{s} \end{pmatrix}$$

by neglecting the "weak" interaction terms. Assuming the cost criterion is formulated in terms of the decoupled systems, as for example 4.3.1, gains k are fed back from the appropriate outputs as shown in Figure 4.3.3, giving

$$J = J_1 + J_2 = \frac{2}{k} .$$

But $$\delta W = \begin{pmatrix} 0 & \frac{-0.1}{0.1s + 1} \\ \frac{-0.1}{0.1s + 1} & 0 \end{pmatrix}$$

$$\therefore \qquad \delta J = \frac{2 \times 0.1}{1 + 0.1k}$$

by evaluating a complex integral as for example 4.3.1.

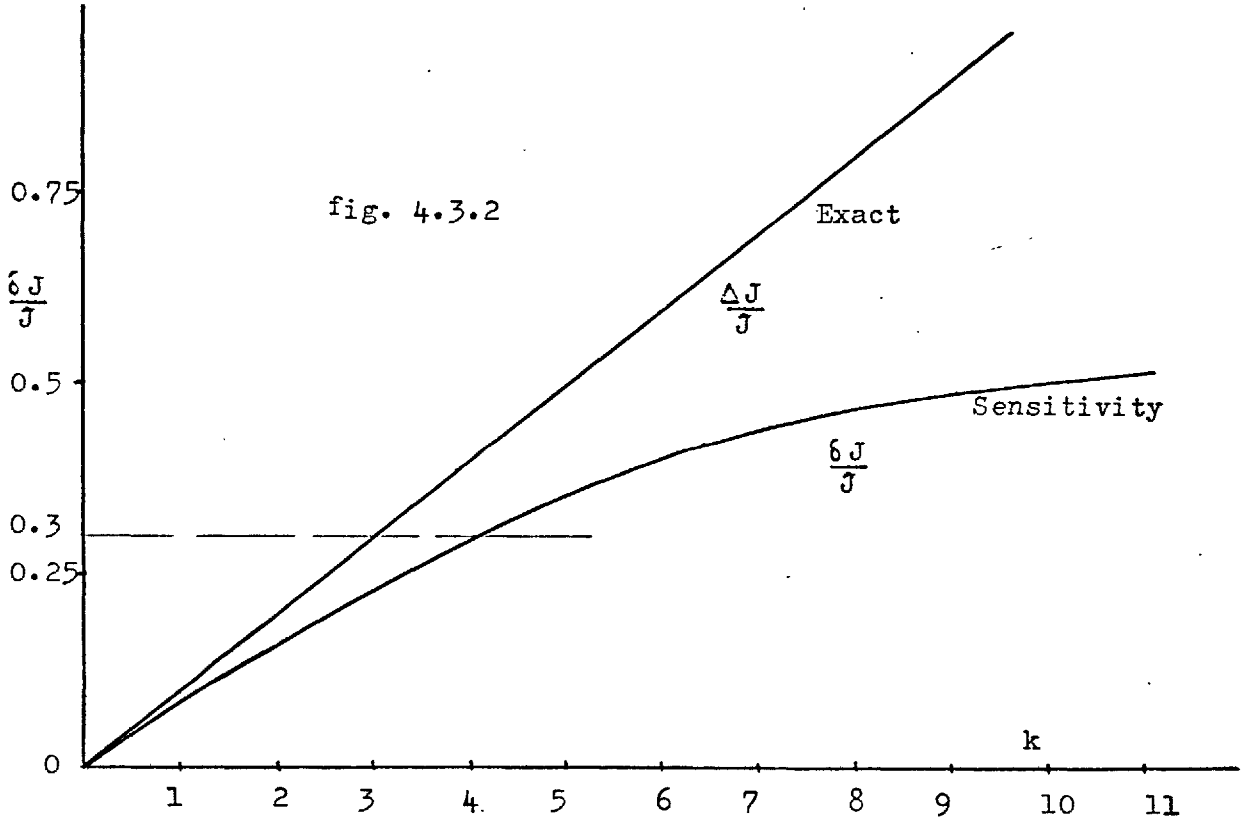


fig. 4.3.2

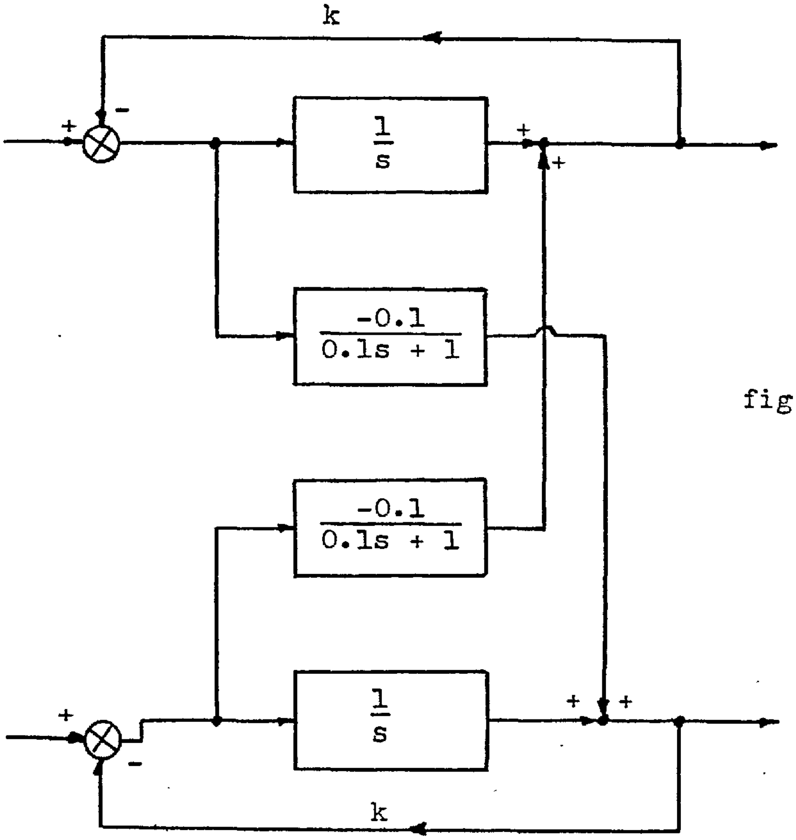


fig.4.3.3

$$\frac{|\delta J|}{J} = \frac{0.1k}{1 + 0.1k} .$$

As distinct from Example 4.3.1, this system may become unstable if k is too large.

Sensitivity analysis indicates what part of the error in weighting function is important, in terms of contribution to δJ. A plot of the cross correlation of y and u gives this immediately. Note that it is the nominal closed loop output and control that determine this sensitivity.

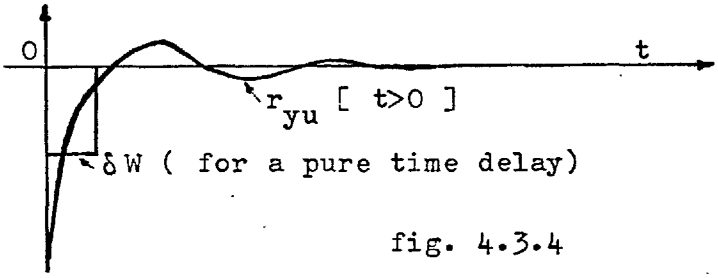


fig. 4.3.4

4.4 Cost Bounds

The previous section has shown how a first order perturbation idea can be utilised to obtain an estimate of the cost change due to a change in system operator. Unfortunately, sensitivity analysis cannot predict system instability, and in many cases it may be desirable to have an absolute bound on the cost change. This is linked with stability, since unstable systems will not have finite cost.

There are two types of bounds that are of use in control design.

(1) Given a sub-optimal control system, we may wish to form an estimate of the difference between the actual cost and the optimal cost. A bound on this difference gives a lower bound to the optimal performance index.

(2) Given an optimal design for an approximate model, a bound on the cost difference between the exact system and the model may also be of use. Sensitivity analysis tried to estimate this difference. Any further refinement, however, has to take into account the neglected second order terms, which may become quite complicated. In fact, the simplest method of obtaining the cost difference is to actually compute the two costs by simulation, and subtract them.

The lower bound to the optimal performance index is more useful in design. One can either bound the difference between the actual cost and the optimal, or find a bound on the ratio.

Theorem 4.4.1: With the system represented by

$$y = y_o + W u$$

and cost $$J = \langle y, Q y \rangle + \langle u, R u \rangle$$

let $$g = R\,u + W^{*} Q\, y\ .$$

Then, if $$A = (R + W^{*} Q\, W)$$

$$\Delta J = J - \hat{J}\ , \text{ where } \hat{J} \text{ is the optimal cost}$$

$$= \langle g, A^{-1} g \rangle\ . \qquad 4.4.1$$

Proof: Note, for $J = \hat{J}$, that

$g = 0$, which is consistent with equation 4.4.1.

Now $$J = \langle y, Q\, y \rangle + \langle u, R\, u \rangle$$

$$= \langle y_o, Q\, y \rangle + \langle W\, u, Q\, y \rangle + \langle u, R\, u \rangle$$

$$= \langle y_o, Q\, y \rangle + \langle u, W^{*} Q\, y + R\, u \rangle$$

$$= \langle y_o, Q\, y \rangle + \langle u, g \rangle\ .$$

$$g = R\,u + W^{*} Q\, y$$

$$= R\,u + W^{*} Q\, (y_o + W\,u)$$

$$= (R + W^{*} Q\, W)u + W^{*} Q\, y_o$$

$$= A\,u + W^{*} Q\, y_o\ . \qquad 4.4.2$$

However $$y = \hat{y} + W(u - \hat{u})\ .$$

So $$J = \langle y_o, Q\, \hat{y} \rangle + \langle y_o; Q\, W(u - \hat{u}) \rangle + \langle u, g \rangle\ .$$

From theorem 2.2.3,

$$\hat{J} = \langle y_o , Q\, y \rangle .$$

Hence $$J = \hat{J} + \langle y_o , Q\, W(u - \hat{u}) \rangle + \langle u, g \rangle$$

$$= \hat{J} + \langle W^* Q\, y_o , (u - \hat{u}) \rangle + \langle u, g \rangle .$$

Since $\hat{u}$ is optimal,

$$A\, \hat{u} = - W^* Q\, y_o \quad .$$

Also $$A\, u = g - W^* Q\, y_o \quad .$$

Since A is self-adjoint and bounded below, A^{-1} exists, and

$$u = A^{-1} g - A^{-1} W^* Q\, y_o \quad .$$

i.e. $$u - \hat{u} = A^{-1} g \quad .$$

Hence $$J = \hat{J} + \langle W^* Q\, y_o , A^{-1} g \rangle + \langle A^{-1} g, g \rangle$$

$$- \langle W^* Q\, y_o , A^{-1} g \rangle$$

$$= \hat{J} + \langle g, A^{-1} g \rangle .$$

<u>Corollary 4.4.1</u>: (Allwright [A 3])

If $$m \langle u, u \rangle \leq \langle u, R\, u \rangle , \qquad 4.4.3$$

then $$\delta J \leq \frac{1}{m} \langle g, g \rangle . \qquad 4.4.4$$

<u>Proof</u>: $\delta J = \langle g, A^{-1} g \rangle$

where $A = R + W^{*} Q W$.

Now $\langle u, (R + W^{*} Q W)u \rangle \geq \langle u, R u \rangle$ since $W^{*} Q W$ is P.S.D.

$\geq m \langle u, u \rangle$.

Let T^2 be the factorisation of A, such that T^{-1} exists, and T is self-adjoint.

i.e. $T^2 = R + W^{*} Q W$.

Then $\langle T u, T u \rangle \geq m \langle u, u \rangle$.

Put $T u = g$ $g = Ru + W^{*}Qy$

i.e. $u = T^{-1} g$.

Then $\langle g, g \rangle \geq m \langle T^{-1} g, T^{-1} g \rangle$

$= m \langle g \, T^{-1} T^{-1} g \rangle$

$= m \langle g, A^{-1} g \rangle$.

$\therefore \quad \langle g, A^{-1} g \rangle \leq \frac{1}{m} \langle g, g \rangle$.

This bound can be quite good when the system is near optimal, but rather poor away from the optimum. It has the desirable property of getting better as the optimum is approached. Allwright [A 3] has used this to good effect to determine stopping criteria for optimisation problems. We envisage its use in the investigation of near optimal approximate designs. If W is an approximate transfer function, and $W + \delta W$ the exact transfer, then

$$g = R(u + \delta u) + (W^* + \delta W^*)Q(y + \delta y) .$$

If u, y represent the nominal optimal trajectories, then

$$\begin{aligned} g &= (R\,u + W^*\,Q\,y) + R\,\delta u + W^*\,Q\,\delta y + \delta W^*\,Q\,y + \delta W^*\,Q\,\delta y \\ &= R\,\delta u + (W + \delta W)^*\,Q\,\delta y + \delta W^*\,Q\,y .\end{aligned}$$

The lower bound may be evaluated quite conveniently, and since m is independent of the dynamics, no factorisation or inversion of operators is required.

An upper bound on the ratio $\frac{J_o}{J}$ can be a useful engineering parameter. J_o is the uncontrolled cost. Theorem 2.2.3 can be used to obtain a result originally due to Brockett.

Theorem 4.4.2:

If $\quad J_o = \langle y_o, Q\,y_o \rangle \qquad$ 4.4.5

and $\quad \hat{J} = \langle y, Q\,y \rangle + \langle u, R\,u \rangle$

where u is an optimal control, then

$$\frac{J_o}{\hat{J}} \leq 1 + \|Q\|\,\|R^{-1}\|\,\|W^*W\| . \qquad 4.4.6$$

Proof: From Theorem 2.2.3,

$$\begin{aligned}\hat{J} &= \langle y_o, Q\, y \rangle \\ &= \langle y_o, Q(y_o + W\, u)\rangle \\ &= J_o + \langle y_o, Q\, W\, u\rangle .\end{aligned}$$

But $u = - R^{-1} W^* Q\, y$.

$$\therefore \quad \hat{J} = J_o - \langle y_o, Q\, W\, R^{-1} W^* Q\, y\rangle .$$

$$\begin{aligned}J_o - \hat{J} &= \langle y_o, Q\, W\, R^{-1} W^* Q\, y\rangle \\ &= \langle W^* Q\, y_o, R^{-1} W^* Q\, y\rangle \\ &= \langle W^* Q\, y_o + W^* Q\, W\, u - W^* Q\, W\, u, R^{-1} W^* Q\, y\rangle \\ &= \langle W^* Q\, y, R^{-1} W^* Q\, y\rangle + \langle W^* Q\, W\, u, u\rangle .\end{aligned}$$

Let $S^* S = Q$, and let $R^{\frac{1}{2}}$ be the P.D. square root of R. Then

$$\begin{aligned}J_o - \hat{J} &= \langle S\, y, (S\, W\, R^{-1} W^* S^*) S\, y\rangle + \langle W\, R^{-\frac{1}{2}} R^{\frac{1}{2}} u, Q\, W\, R^{-\frac{1}{2}} R^{\frac{1}{2}} u\rangle \\ &\leq \| S\, W\, R^{-1} W^* S^* \| \langle S\, y, S\, y\rangle + \|R^{-\frac{1}{2}} \; W^* Q\, W\, R^{-\frac{1}{2}}\| \langle R^{\frac{1}{2}} u, R^{\frac{1}{2}} u \\ &\leq \|S\| . \|S^*\| . \|R^{-1}\| . \|W\| . \|W^*\| \langle y, Q\, y\rangle \\ &\qquad + \|R^{-\frac{1}{2}}\| \, \| R^{-\frac{1}{2}} \| \, \|Q\| \, \| W^* \| \langle u, R\, u\rangle \\ &= \|Q\| \, \|R^{-1}\| \, \| W^* W \| (\langle y, Q\, y\rangle + \langle u, R\, u\rangle) .\end{aligned}$$

Since $\qquad \|W\| = \|W^*\| = \|W^* W\|^{\frac{1}{2}} = \|W W^*\|^{\frac{1}{2}}$

(see lemma 4.5.1),

$$J_o - \hat{J} \leq (\|Q\| \, \|R^{-1}\| \, \|W^* W\|) \, \hat{J}$$

$$\frac{J_o}{\hat{J}} \leq 1 + \|Q\| \, \|R^{-1}\| \, \|W^* W\| .$$

This is a convenient bound for estimating the decrease in cost expected by performing control action, but is not too useful for determining whether a given closed loop system approximates optimal control. We have not been successful in deriving a convenient bound of this kind. It is felt that a possible starting point would be to use the following lemma, which is a generalisation of Theorem 2.2.3.

<u>Lemma 4.4.1</u>: If y, u are the sub-optimal outputs and controls respectively, and $\hat{y}, \hat{u}$ are optimal, then

$$\hat{J} = \langle y, Q\hat{y} \rangle + \langle u, R\hat{u} \rangle . \qquad 4.4.7$$

<u>Proof</u>: $\qquad J = \langle y_o, Q\hat{y} \rangle \qquad$ from theorem 2.2.3,

$$= \langle y - Wu, Q\hat{y} \rangle$$

$$= \langle y, Q\hat{y} \rangle - \langle Wu, Q\hat{y} \rangle$$

$$= \langle y, Q\hat{y} \rangle - \langle u, W^* Q\hat{y} \rangle .$$

But for optimality,

$$R\hat{u} + W^* Q\hat{y} = 0 .$$

Hence $\hat{J} = \langle y, Q\,\hat{y} \rangle + \langle u, R\,\hat{u} \rangle .$

The continuation of the derivation similar to that of theorem 4.4.2 becomes very messy.

4.5 Norms and Stability

Operator norms have been freely used throughout this thesis. The basic definition is given in terms of the associated vector norm:

$$\| W \| = \sup_{u \neq o} \frac{\| W u \|}{\| u \|} . \qquad 4.5.1$$

If $u \in \mathcal{H}$, a Hilbert space, and the norm of u is defined in terms of the inner product, then

$$\| u \|^2 = \langle u, u \rangle . \qquad 4.5.2$$

Using this vector norm, we obtain the following:

Lemma 4.5.1

$$\| W \| = \| W^{*} \| = \| W W^{*} \|^{\frac{1}{2}} = \| W^{*} W \|^{\frac{1}{2}} . \qquad 4.5.3$$

Proof: $$\| W \| = \sup_{u \neq o} \left(\frac{\langle W u, W u \rangle}{\langle u, u \rangle} \right)^{\frac{1}{2}}$$

i.e. $$\| W \|^2 = \sup_{u \neq o} \frac{\langle W u, W u \rangle}{\langle u, u \rangle}$$

$$= \sup_{u \neq o} \frac{< u, W^* W u >}{< u, u >}$$

But $$\frac{<u, W W^* u>}{\|u\|^2} \leq \frac{\|u\| \, \|W^* W u\|}{\|u\|^2}$$

by the Schwartz inequality, and this bound is attained.

i.e. $$\|W\|^2 = \sup_{u \neq o} \frac{\|W^* W u\|}{\|u\|}$$

$$= \|W^* W\| .$$

From theorem 1.5.1, $\|W\| = \|W^*\|$, and using the above analysis with W^* replacing W, the result is obtained.

For time-varying continuous weighting functions over a finite interval, the derivation of Section 2.4 shows that, for single-input/single-output systems

$$\|W\|^2 \leq \int_o^T (\int_o^t | W(t, \tau)|^2 \, d\tau) dt . \qquad 4.5.4$$

This is, in general, a very crude bound, and usually is infinite for infinite intervals. What is really necessary is to evaluate the maximum eigenvalue of $W^* W$, and take its square root. However, this can be quite difficult, though iterative techniques are available.

For time invariant systems over an infinite interval, the following analysis produces better bounds to the norm. Taking transforms,

$$y(s) = W(s)\,u(s) .$$

$$\therefore \qquad \|y\|^2 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} u^T(-s)\,W^T(-s)\,u(s)\,ds$$

$$\leqq \sup_{\omega}\ \lambda_{max}[W^T(-j\omega)\,W(j\omega)]\ \|u\|^2 \qquad - \qquad 4.5.5$$

$$\leqq \sup_{\omega}\ tr[W^T(-j\omega)\,W(j\omega)]\,\|u\|^2 . \qquad 4.5.6$$

A result in the time domain for multivariable systems can be obtained from equation 1.7.4. If subscripts denote vector coordinates, then

$$y_k(t) = \int_o^t \sum_i W_{ki}(t-\tau) . u_i(\tau)\,d\tau$$

$$= \sum_i \int_o^t W_{ki}(t-\tau) . u_i(\tau)\,d\tau .$$

From 1.7.4 $\|y_k\|_2 \leqq (\sum_i \|W_{ki}\|_1 \|u_i\|_2)$. 4.5.7

But $\|y\|_2^2 = \sum_k \|y_k\|_2^2$.

Let $\|W_{ki}\|_1 = a_{ki}$, the coefficients of matrix A,

$\|y_k\|_2 = c_k$, the coefficients of a vector $\bar{c}$,

$\|u_i\|_2 = b_i$, the coefficients of a vector b.

Then, if $\quad d = A\,b \,, \quad$ where $\quad d = (d_k)$

$$|c_k| \leq |d_k| \qquad \text{for each } k .$$

i.e. $\quad ||c|| \leq ||d|| \,,$

$$\leq (\sqrt{\lambda_{max}\,(A^T\,A)}\;||b|| \,. \qquad 4.5.8$$

i.e. $(\lambda_{max}\,A^T\,A)^{\frac{1}{2}}$ is a bound on the norm of W, where

$$a_{ij} = \int_0^\infty |W_{ij}(t)|\,dt \,.$$

It may be simpler to calculate the trace of $A^T\,A$, and then use

$$(\lambda_{max}\,A^T\,A)^{\frac{1}{2}} \leq (tr\;A^T\,A)^{\frac{1}{2}} \,. \qquad 4.5.9$$

Consider an operator G, which is used to form a feedback system, with unity feedback, so that

$$F = (I + G)^{-1}\,G = G(I + G)^{-1} \qquad 4.5.10$$

where F is the closed-loop transfer operator, which we assume stable; i.e. $||\,F\,|| < \infty$. If there is a perturbation δG in G, it is of interest to find a bound on the norm of δF.

Theorem 4.5.2:

$$||\delta F|| \leq \frac{||\,(I + G)^{-1}\,\delta G(I + G)^{-1}\,||}{1 - ||(I + G)^{-1}\,\delta G\,||} \qquad 4.5.11$$

providing $\qquad ||\,(I + G)^{-1}\,\delta G\,|| < 1 \qquad 4.5.12$

<u>Proof</u>: $$F = (I + G)^{-1} G = G(I + G)^{-1} .$$

i.e. $$F + GF = G .$$

$\therefore$ $$(F + \delta F) + (G + \delta G)(F + \delta F) = G + \delta G .$$

i.e. $$F + \delta F + GF + \delta . GF + G.\delta F + \delta G.\delta F = G + \delta G$$

$$\delta F + G.\delta F + \delta . GF + \delta G.\delta F = \delta G$$

$$(I + G)\delta F = \delta G - \delta G.F - \delta G.\delta F$$

$$= \delta G(I - F) - \delta G.\delta F .$$

From 4.5.10, $$I - F = I - (I + G)^{-1}G$$

$$= (I + G)^{-1}(I + G - G)$$

$$= (I + G)^{-1} .$$

$\therefore$ $$\delta F = (I + G)^{-1}\delta G(I + G)^{-1} - (I + G)^{-1}\delta G.\delta F$$

$$\|\delta F\| = \| (I + G)^{-1}\delta G(I + G)^{-1} - (I + G)^{-1}\delta G.\delta F \|$$

$$\leq \|(I + G)^{-1}\delta G(I + G)^{-1}\| + \|(I + G)^{-1}\delta G\| \, \|\delta F\|$$

i.e. $$\|\delta F\| \leq \frac{\|(I + G)^{-1}\delta G(I + G)^{-1}\|}{1 - \|(I + G)^{-1} \delta G \|}$$

providing $\|(I + G)^{-1}\delta G \| < 1$.

<u>Lemma 4.5.2</u>: If $m^2 \langle u, u\rangle \leq \langle T u, T u\rangle$

Then
$$\| T^{-1}\| \leq \frac{1}{m} \qquad 4.5.13$$

<u>Proof</u>: Since T is bounded below, T^{-1} exists.

But
$$\| T^{-1}\| = \sup \frac{\| T^{-1}y\|}{\| y \|}$$

Put
$$u = T^{-1}y$$

$$y = T u \quad .$$

But
$$m^2 \langle T^{-1}y, T^{-1}y \rangle \leq \langle y, y \rangle$$

$$\frac{1}{m^2} \geqq \frac{\langle T^{-1}y, T^{-1}y \rangle}{\langle y, y\rangle}$$

i.e.
$$\| T^{-1} \| \leq \frac{1}{m} \; .$$

<u>Theorem 4.5.3</u>: If

$$(I + G)^*(I + G) = I + W^* Q W \qquad 4.5.14$$

is the spectral factorisation solution to an optimal control problem,

then
$$\| (I + G)^{-1}\| \leqq 1 \quad .$$

<u>Proof</u>: $\langle u,(I + G)^*(I + G)u\rangle = \langle u, (I + W^* Q W)u \rangle$

$$\geqq \langle u, u \rangle \quad .$$

Hence, by lemma 4.5.2,

$$\| (I + G)^{-1} \| \leqq 1 \ . \qquad 4.5.15$$

For G satisfying 4.5.15, equation 4.5.11 becomes

$$\begin{aligned} \|\delta F\| &\leqq \frac{\|(I + G)^{-1}\delta G(I + G)^{-1}\|}{1 - \|(I + G)^{-1}\delta G\|} \\ &\leqq \frac{\|(I + G)^{-1}\|^2 \|\delta G\|}{1 - \|(I + G)^{-1}\| \|\delta G\|} \\ &\leqq \frac{\|\delta G\|}{1 - \|\delta G\|} \ . \qquad 4.5.16 \end{aligned}$$

In general, this bound is coarser than that of equation 4.5.11, but is in general easier to calculate if equation 4.5.15 is applicable. It is quite good when δG is large at frequencies where G is small, as is shown by the following simple example.

<u>Example 4.5.1</u>:

Complex model: $W(s) = \frac{e^{-sT}}{s}$

Simple model: $W_1(s) = \frac{1}{s}$.

For an optimal design using the simple model

$$G(s) = \frac{k}{s} ,$$

where $$J = \int_0^\infty y^2 + \frac{1}{k^2} u^2 \, dt \, .$$

Hence $$\delta G(s) = (\frac{1 - e^{-sT}}{s}) \, k \, .$$

$$\delta g(t) = \mathcal{L}^{-1}(\delta G) = k \quad \text{for} \quad 0 \leq t \leq T$$

and zero otherwise.

$$\delta G(-j\omega)\delta G(j\omega) = \frac{k^2}{\omega^2} \, 2(1 - \cos \omega T)$$

$$= \frac{(kT)^2}{(\omega T)^2} \, 2(1 - \cos \omega T)$$

Put $x = kT$ and $\theta = \omega T$.

Then $$|\delta G|^2 = \frac{x^2}{\theta^2} \, 2(1 - \cos \theta) \, .$$

This achieves a maximum for

$$\theta \sin \theta = 2(1 - \cos \theta)$$

i.e. $$\theta = 0 \, .$$

$$\therefore \qquad \delta G = x = kT = \| \delta g \|_1 \, .$$

Instability, on the basis of equation 4.5.16, is predicted when

$$kT = 1 \, .$$

From exact analysis [C2]

$$kT = \frac{\pi}{2} = 1.57$$

is the stability limit.

A better bound is achieved by considering the value of kT for which

$$\left\| \frac{\delta G}{1 + G} \right\| = 1 .$$

But
$$\left| \frac{\delta G}{1 + G} \right|^2 = k^2 \left| \frac{1 - e^{-sT}}{s + k} \right|^2$$

$$= \frac{2k^2(1 - \cos \omega T)}{k^2 - \omega^2} \qquad \text{for } s = j\omega$$

$$= \frac{2x^2(1 - \cos \theta)}{x^2 + \theta^2} .$$

This achieves a maximum for

$$(x^2 + \theta^2) \sin \theta = 2\theta(1 - \cos \theta)$$

i.e.
$$\frac{\sin \theta}{2\theta} = \frac{1 - \cos \theta}{x^2 + \theta^2} .$$

But
$$\left| \frac{\delta G}{1 + G} \right|^2 = 1 = \frac{x^2 \sin \theta}{\theta} \quad \text{for stability limit.}$$

Substituting for x^2 gives

$$\theta + \theta^2 \sin \theta = 2\theta(1 - \cos \theta)$$

$$\theta \sin \theta = 1 - 2 \cos \theta .$$

From tables $\theta \sim 2$ radians .

i.e. $x^2 = \frac{2}{0.9} = 2.22$.

$\therefore$ $kT = 2.22 = 1.45$.

This is only slightly below the exact stability limit of 1.57.

Equation 4.5.16 has a simple interpretation on the Nyquist diagram (Figure 4.5.1).

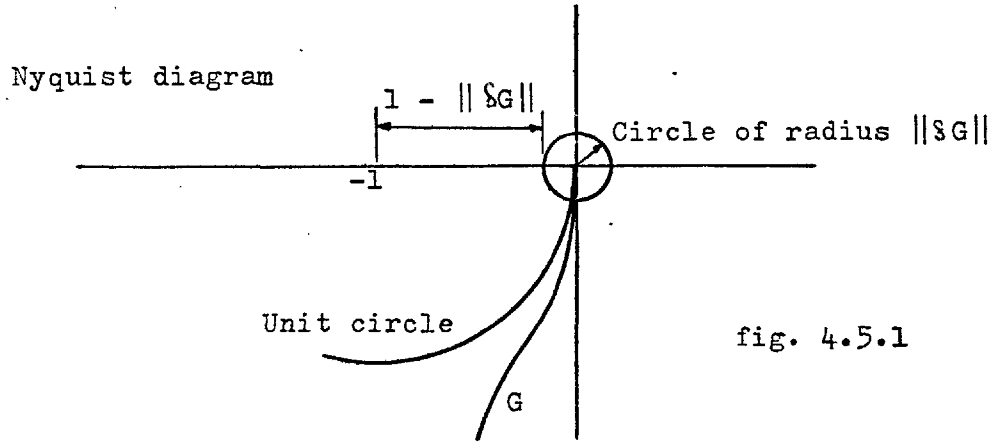


fig. 4.5.1

An error of given norm $\|\delta G\|$ produces its worst effect at $j\omega = \infty$. In this case, instability may occur for $1 - \|\delta G\|$ zero.

4.6 "A Priori" Design Methods

The methods discussed so far for determining the validity of approximate design methods have all be "a posteriori" methods. That is, a check is performed after the control system has been designed, although it may be possible to keep a variable parameter in the design, such as a feedback gain, to adjust during the checking stage. If design entails a large amount of work, it may be preferable to have some "a priori" guide to the selection of design parameters, chosen on the basis of model uncertainty. In particular, for optimal control law calculations, guides to the choice of Q and R are useful.

Due to fundamental limitations in measurement, the impulse response of a system can never be exactly known, even assuming linearity. Hence it seems reasonable to model an impulse response $W_1(t)$ as a non-stationary stochastic process:

$$W_1(t) = W(t) + N(t) \qquad 4.6.1$$

where

$$E(W_1(t)) = W(t) \qquad 4.6.2$$

and

$$E(N(t)) = 0 . \qquad 4.6.3$$

We shall further assume that $N(t)$ is a white noise process, but with time-varying variance.

$$E\, N^T(t)\, Q\, N(\tau) = R_1(t)\, \delta(t - \tau) \qquad 4.6.4$$

where $R_1(t)$ is positive semi-definite, and integrable over $[0,\infty)$ such that

$$R_2 = \int_0^\infty R_1(t)\, dt \qquad 4.6.5$$

is positive definite. Thus, the error in modelling is represented by a decaying white noise burst centred on the model as mean, as shown in Figure 4.6.1.

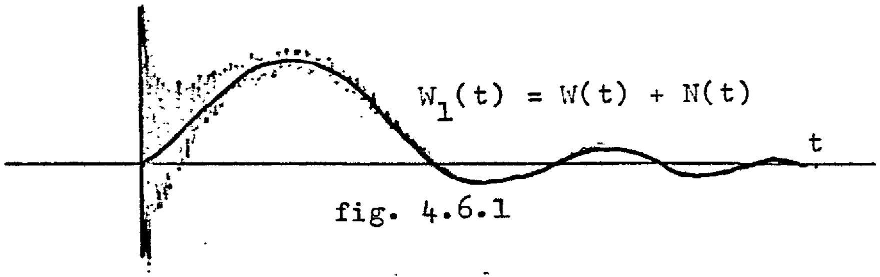


fig. 4.6.1

Let
$$y_1(t) = y_o(t) + \int_0^t W_1(t-\tau)\, u(\tau)\, d\tau \qquad 4.6.6$$

$$= y_o(t) + \int_0^t W(t-\tau)\, u(\tau)\, d\tau + \int_0^t N(t-\tau)\, u(\tau)\, d\tau$$

$$= y(t) + \int_0^t N(t-\tau)\, u(\tau)\, d\tau \qquad 4.6.7$$

$$= y(t) + v(t) \;. \qquad 4.6.8$$

We wish to find a determinstic $u(t)$ to minimise the cost function

$$J = E \int_0^\infty y_1^T(t)\, Q\, y_1(t)\, dt + \int_0^\infty u^T R\, u\, dt \qquad 4.6.9$$

where Q is chosen symmetric, and positive definite, to weight the relative importance of the outputs and their interaction. However,

$$J = E \int_0^\infty (y(t) + v(t))^T Q (y(t) + v(t)) dt + \int_0^\infty u^T R\, u\, dt$$

$$= E \int_0^\infty (y^T Q\, y + 2y^T Q\, v + v^T Q\, v) dt + \int_0^\infty u^T R\, u\, dt\, . \qquad 4.6.10$$

If u is deterministic, so is y. Hence

$$J = \int_0^\infty y^T Q\, y\, dt + \int_0^\infty 2y^T Q\; E(v)\; dt + E \int_0^\infty v^T Q\, v\, dt$$

$$+ \int_0^\infty u^T R\, u\, dt\, . \qquad 4.6.11$$

But

$$E(v(t)) = E \int_0^t N(\tau)\, u(t - \tau)\, d\tau$$

$$= \int_0^\infty E\, N(\tau)\;\; u(t - \tau)\, d\tau$$

$$= 0 \quad . \qquad 4.6.12$$

$$E \int_0^\infty v^T Q\, v\, dt = \int_0^\infty E(v^T Q\, v)\, dt\, . \qquad 4.6.13$$

$$E(v^T Q v) = E[\int_0^t u^T(\tau) N^T(t-\tau)\, d\tau \,.\, Q \int_0^t N(t-s)\, u(s)\, ds]$$

$$= \int_0^t d\tau \int_0^t ds\, u^T(\tau)\, [E[N^T(t-\tau)\, Q\, N(t-s)\,]]\, u(s). \qquad 4.6.14$$

From 4.6.4

$$E[N^T(t-\tau)\, Q\, N(t-s)] = R_1(t-\tau)\, \delta(\tau - s) \quad . \qquad 4.6.15$$

$$\text{So} \quad E(v^T Q v) = \int_0^t u^T(\tau)\, (\int_0^t R_1(t-\tau)\, \delta(\tau-s)\, u(s)\, ds)\, d\tau$$

$$= \int_0^t u^T(\tau)\, R_1(t-\tau)\, u(\tau)\, d\tau \quad . \qquad 4.6.16$$

$$\therefore \int_0^\infty E(v^T Q v)\, dt = \int_0^\infty \int_0^t u^T(\tau)\, R_1(t-\tau)\, u(\tau)\, d\tau \;\; dt. \qquad 4.6.17$$

Changing the order of integration (which we assume permissible),

$$\int_0^\infty E(v^T Q v)\, dt = \int_0^\infty [\int_\tau^\infty u^T(\tau)\, R_1(t-\tau)\, u(\tau)\, dt\,]\, d\tau$$

$$= \int_0^\infty u^T(t)\, [\int_t^\infty R_1(\tau - t)\, d\tau]\, u(t)\, dt, \qquad 4.6.18$$

where the roles of the dummy variables t and τ have been interchanged.

But $$\int_t^{\infty} R_1(\tau - t)\, d\tau = \int_o^{\infty} R_1(s)\, ds$$

$$= R_2 \quad . \qquad 4.6.19$$

i.e. $$J = \int_o^{\infty} y^T Q\, y + u^T(R_2 + R)u\, dt \qquad 4.6.20$$

where $$y(t) = y_o(t) + \int_o^{t} W(t - \tau)\, u(\tau)\, d\tau \quad . \qquad 4.6.21$$

This analysis indicates that the deterministic control u should be designed as though the noise N were not present, although the cost on control should be increased by R_2. We have an explicit solution to the intuitive idea that, if one is not too sure of the effect of control, one should not put too much of it in.

If W represents a simple low-order state model, and N perturbations about this, then u should be generated, ideally, from the states of the model, and not the perturbed states, since otherwise stability is no longer guaranteed. However, it seems reasonable to try feeding back from the inexact outputs, as an approximation to the desired closed loop system.

From theorem 4.4.2, the cost reduction due to optimal control depends on $\| Q \|\, \|R^{-1}\|\, \|W^{*} W\|$, which is a measure of the square of the system gain. The smaller the value of R^{-1} the smaller the allowable system gain. That is, the smaller the errors, the larger the permissible gain. If $R >> R_2$, then R_2 can be neglected and our approximation is valid, even in closed loop operation.

When the modelling error is known as a deterministic quantity, as in system approximation, or model reduction, the application of the above technique is not immediately obvious. We shall use the following heuristic argument. Let

$$\delta G_{ij}(t) = |\delta W_{ij}(t)| \quad . \qquad 4.6.22$$

We shall consider δG as a kind of standard deviation of our dynamic error. But from the stochastic case

$$\int_0^\infty (\int_0^\infty E(N^T(t) \; Q \; N(\tau)) \, d\tau) \, dt$$

$$= \int_0^\infty \int_0^\infty R_1(t) \, \delta(t-\tau) \, d\tau \, dt$$

$$= \int_0^\infty R_1(t) \, (\int_0^\infty \delta(t-\tau) \, d\tau) \, dt$$

$$= \int_0^\infty R_1(t) \, dt = R_2 \, .$$

Applying the same analysis to δG, we obtain

$$R_2 = \int_0^\infty \int_0^\infty E(\delta G^T(t) \; Q \; \delta G(\tau)) \, d\tau \, dt$$

$$= \left(\int_0^\infty \delta G(t)\, dt \right)^T Q \left(\int_0^\infty \delta G(\tau)\, d\tau \right) .$$

This argument is a plausibility argument only, but it seems the correct kind of limitation to impose, as the following simple examples show.

Example 4.6.1: For the models of example 4.5.1,

$$\delta W = \frac{1 - e^{-sT}}{s} .$$

i.e. $\delta W(t) = 1 \qquad$ for $0 \leqq t \leqq T$

$\qquad = 0 \qquad t < 0, \quad t > T$.

Then $R = \left(\int_0^T 1\, dt \right)^2 q$

$$= T^2 q .$$

We now design the control to minimise

$$J = \int_0^\infty q y^2 + T^2 q u^2 \, dt ,$$

with the plant

$$y(s) = \frac{1}{s} (1 + u(s)) .$$

The problem has the simple answer

$$k^2 = \frac{1}{T^2} \qquad \text{where} \qquad u = -k y .$$

$$\therefore \qquad kT = 1$$

which gives a good stability margin, since instability results for $kT = \pi/2$ (see example 4.5.1).

Example 4.6.2:

Complex model: $\frac{1}{s}\left(\frac{a-s}{a+s}\right)$

Simple model: $\frac{1}{s}$

Then $$\delta W(s) = \frac{2}{a+s}$$

$\therefore$ as above $$R = \frac{4}{a^2},$$

and the maximum gain is $k = \frac{a}{2}$, whereas instability results for a gain of $k = a$.

Example 4.6.3: Weak coupling in multivariable systems.

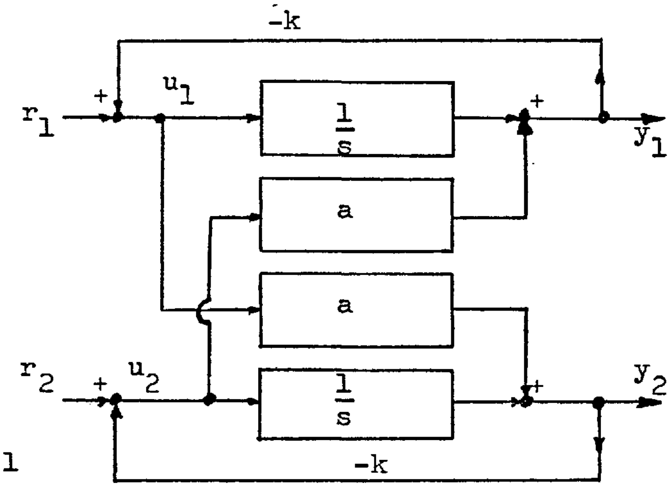


fig. 4.6.1

Consider the system of Figure 4.6.1. The cross-coupling gains a are small and we shall neglect them, and perform an optimal design on the uncoupled simplified model, leading to gains k as shown on the diagram. With this configuration, the closed loop response from r_1 to y_1 is given by

$$\frac{y_1}{r_1} = \frac{\frac{1}{s}}{1 + k(\frac{1}{s} - \frac{a^2 k}{1 + ks})} .$$

The characteristic polynomial $p(s)$ is given by

$$p(s) = s^2 k + s(1 + k^2 - a^2 k^2) + k ,$$

and instability results for $a^2 = (1 + k^2)/k^2$.

i.e. $$k^2 = \frac{1}{a^2 - 1} , \text{ i.e. } |a| > 1 .$$

From the preceding considerations, R is chosen such that

$$R \geqq \begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix} \begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix}$$

$$\geqq \begin{pmatrix} a^2 & 0 \\ 0 & a^2 \end{pmatrix}$$

But $$R = \begin{pmatrix} \frac{1}{k^2} & 0 \\ 0 & \frac{1}{k^2} \end{pmatrix}$$

i.e. $$k^2 a^2 \geqq 1 .$$

This ensures closed loop stability.

The topics and methods of this chapter suggest much further research. While a lot of the argument has been of a heuristic kind, the ideas seem to indicate good methods of performing sub-optimal designs, using optimal control theory as a design technique.

CHAPTER 5

FIXED STRUCTURE OPTIMISATION

5.1 Introduction

Optimal control theory, as formulated for the state-space approach, leads to a feedback control law which specifies the control as a linear combination of all the states. The results of the last chapter, together with engineering experience, indicate that only the "dominant" states need to be fed back (or estimated and fed back). However, "dominance" is a function of the closed loop design, rather than the open loop system. One approach is to approximate the actual system with a state-space model of the appropriate dimension, to give the desired control structure. This has been discussed in Chapter 4. The following approach is the subject of this chapter:

1. On the basis of engineering experience, guided by the dominance concept, a control structure is proposed but with unknown parameters.

2. A quadratic cost function in terms of input and output signals is optimised with respect to these parameters.

The choice of a useful cost function is still open. Note that the cost criterion will not in general be quadratic in the unknown parameters. The limitations of this method will become apparent from the discussion, but there are many advantages.

1. A system of pre-specified complexity is designed and, for a given cost criterion and disturbance, the parameters are the best possible.

2. A feedback design may be achieved, with all the attendant desirable properties such as sensitivity reduction, linearisation, and disturbance rejection.

3. No approximation or modification of the linear model is required before a design is attempted.

There are some disadvantages of this technique.

1. In a complicated multivariable system it may not be obvious that a given structure will enable good control to be achieved. The investigation of this subject is still a valid topic for research.

2. The system is, in general, only optimised for one particular disturbance, and its performance under other disturbances is not explicitly known.

The main contributions of this chapter are the presentation of algorithms for optimising parameters in fixed structures. Of course, any general hill-climbing technique could be used for this problem, but the methods that we propose take advantage of the structure of the problem, and can be very efficient. Large numbers of cost evaluations, and line-minimisations are avoided. The original guide to this piece of research was an iterative method for solving the steady-state Riccati equation, which is demonstrated in the next section.

5.2 An Iterative Method for Solving the Steady-State Riccati Equation

The following derivation considers the solution of the continuous time Riccati equation. Similar results hold for the discrete time problem, and these are summarised at the end of this section.

Consider the steady-state equation

$$P A + A^T P + Q - P B R^{-1} B^T P = 0 \qquad 5.2.1$$

where P is positive definite, Q is positive semi-definite, R is positive-definite, and A, B is controllable. We shall also restrict the discussion to the case where A is initially strictly stable; i.e. all eigenvalues of A lie in the open left half complex plane. Then the following well-known theorem can be stated, as a particular result of Lyapunov stability theory.

Theorem 5.2.1: If A is a strictly stable matrix, then given any positive semi-definite matrix Q_1, there exists a symmetric positive definite matrix P_1, such that

$$P_1 A + A^T P_1 = - Q_1 . \qquad 5.2.2$$

Using this fact, the following iterative method of solution of equation 5.2.1 has been proposed [KL 1].

Algorithm 5.2.1

1. At iteration 0, let $A_o = A$, and $P_{-1} = 0$. Alternatively, if A is unstable, choose K_{-1} such that $A_o = A - B K_{-1}$ is stable.

2. Solve

$$P_i A_{i-1} + A_{i-1}^T P_i = - Q - K_{i-1}^T R K_{i-1} \qquad 5.2.3$$

for P_i positive definite. The existence of such a P_i follows from theorem 5.2.1.

3. Set $$K_i = + R^{-1} B^T P_i \qquad 5.2.4$$

4. Set $$A_i = A - B K_i \qquad 5.2.5$$

5. Return to 2.

Theorem 5.2.2: Algorithm 5.2.1 converges to a solution of equation 5.2.1, where P is positive definite. Moreover, for any $x \neq o$,

$$x^T P_i x < x^T P_{i-1} x \quad , \qquad i = 1, 2, \ldots$$

Proof: [KL 1]. It is easily verified that a fixed point of the algorithm solves equation 5.2.1. The monotone cost decrease ensures convergence.

The corresponding algorithm for discrete time systems is

<u>Algorithm 5.2.2:</u>

$$P_{i+1} = Q + K_i^T R K_i + (A - B K_i)^T P_i (A - B K_i) \qquad 5.2.6$$

$$K_{i+1} = (R + B^T P_i B)^{-1} B^T P_i A \quad . \qquad 5.2.7$$

For numerical work it is not necessary to calculate the P matrix explicitly, if it is only desired to calculate an optimal gain K. For (in the continuous case)

$$P = \int_0^\infty \Phi^T (Q + P B R^{-1} B^T P) \Phi \, dt \qquad 5.2.8$$

where $\Phi = e^{(A - B K)t}$. 5.2.9

Hence $K_{i+1} = R^{-1} B^T (\int_0^\infty \Phi_i^T (Q + K_i^T R K_i) \Phi_i \, dt)$. 5.2.10

Using equation 5.2.10, or a similar one for discrete time, K_{i+1} can be calculated as a functional involving K_i. This is precisely what is done in the following sections, but the rule is generalised to work for non-optimal structures. We first give a simple example of the use of continuous time algorithm, to demonstrate its power.

<u>Example 5.2.1:</u> Continuous time - one state.

$$\dot{x} = a x + u$$

$$J = \int_{o}^{\infty} x^2 + u^2 \, dt .$$

Then $\quad 2(a - k_i)p_{i+1} = -(1 + k_i^2) .$

$$p_{i+1} = -\frac{(1 + k_i^2)}{2(a - k_i)} .$$

But $\quad k = p .$

Therefore $\quad k_{i+1} = \frac{1}{2}\left(\frac{1 + k_i^2}{k_i - a}\right) .$

For $a = -1$, the optimal answer is $k = \sqrt{2} - 1$ (from direct solution of equation 5.2.1). Application of the iteration technique yields the following sequence

$$k_o = 0$$

$$k_1 = 0.5$$

$$k_2 = 0.416666$$

$$k_3 = 0.414214$$

$k_\infty = k_3$ to 6 decimal places !

5.3 Restricted Operators

We digress to introduce some new notation, which enables our equations to be written concisely.

Consider two Hilbert spaces $\mathcal{H}_1$, $\mathcal{H}_2$, where

$$a \in \mathcal{H}_1 \qquad \text{and} \qquad b \in \mathcal{H}_2 \, .$$

Let $K \in \mathcal{J}_{21}$, the space of mappings from $\mathcal{H}_2$ to $\mathcal{H}_1$.

Then $$f = \langle a, K b \rangle$$

is an inner product on $\mathcal{H}_1$. From the previous discussion, this induces a functional on ~~K~~ $\mathcal{J}_{21}$, such that

$$\langle a b^*, K \rangle = \langle a, K b \rangle \, .$$

We give these functionals on $\mathcal{J}_{21}$ a special symbol "tr" which is interpreted as a generalised trace [DS 2]. For the special case that $\mathcal{H}_1$ and $\mathcal{H}_2$ are finite dimensional euclidean spaces, the "tr" operation becomes the normal trace of a square matrix.

$$\begin{aligned} \langle a b^*, K \rangle &= \operatorname{tr} (a b^*)^* K \\ &= \operatorname{tr} b a^* . K \\ &= \operatorname{tr} K . b a^* \, . \end{aligned}$$

We now develop this idea one stage further. Let $\mathcal{H}_y$, $\mathcal{H}_u$, and $\mathcal{H}_t$

be Hilbert spaces, with

$$\mathcal{H}_y = \mathcal{H}_1 \otimes \mathcal{H}_t$$

and
$$\mathcal{H}_u = \mathcal{H}_2 \otimes \mathcal{H}_t \, .$$

The operator $K \in \mathcal{J}_{21}$ induces a map $K \otimes I_t$ from $\mathcal{H}_y$ into $\mathcal{H}_u$, where I_t is the unit operator on $\mathcal{H}_t$. [H 2; p.365-369]. Usually, this operation is also denoted simply by K. Then the inner product

$$\alpha = \langle u, (K \otimes I_t) \, y \rangle$$

is defined. This now induces a bilinear map from $\mathcal{H}_u \oplus \mathcal{H}_y$ into $\mathcal{J}_{21}{}^*$, denoted by $[u, y]$, where

$$\begin{aligned} \langle u, (K \otimes I_t) y \rangle &= \langle [u, y], K \rangle \\ &= \operatorname{tr} \, [u, y]^* K \, . \end{aligned}$$

The [,] operation is a kind of contraction in the sense of section 1.2.

<u>Example 5.3.1</u>: Let $\mathcal{H}_1 = R^n$, $\mathcal{H}_2 = R^m$, and $\mathcal{H}_t = L_2[0, \infty)$. Then if $y \in \mathcal{H}_y$, and $u \in \mathcal{H}_u$

$$\langle u, K y \rangle = \int_0^\infty u^T(t) \, K \, y(t) \, dt \, .$$

K is represented by an $m \times n$ matrix. Then

$$< u, K y > = \int_0^\infty (u^T(t) K) y(t) dt$$

$$= tr \int_0^\infty y(t) u^T(t) K dt$$

$$= tr (\int_0^\infty u(t) y^T(t) dt)^T K .$$

i.e. $$[u, y] = \int_0^\infty u(t) y^T(t) dt .$$

Example 5.3.2: Let $\mathcal{H}_1 = L_2[0, 1]$, $\mathcal{H}_2 = R^1$, and $\mathcal{H}_t = L_2[o, \infty)$ Then K is a functional on $L_2[0, 1]$. If K is represented by a continuous kernel, then

$$< u, K y > = \int_0^\infty u(t) (\int_0^1 K(s) y(s, t) ds) dt$$

$$= \int_0^\infty \int_0^1 u(t) K(s) y(s, t) ds dt$$

$$= \int_0^1 (\int_0^\infty u(t) y(s,t) dt) K(s) ds .$$

$$[u, y] = \int_0^\infty u(t) y(s, t) dt .$$

Example 5.3.3: $\mathcal{H}_1 = R^n$, $\mathcal{H}_2 = R^m$, $\mathcal{H}_t$ = space of one-dimensional stationary stochastic processes with zero mean and finite variance.

$$< u, K\, y > \; = \; E(u^T(t)\, K\, y(t))$$

$$= \; \mathrm{tr}[\, E(u(t)\, y^T(t))]^T\, K \quad . \quad //$$

If $M, K \in \mathcal{J}_{21}$, then $< M\, y_1, K\, y_2 >$ is an inner product on $\mathcal{H}_u$, where $y_1, y_2 \in \mathcal{H}_y$.

Lemma 5.3.1:

$$< [\, M\, y_1, y_2], K > \; = \; < M[\, y_1, y_2], K>$$

Proof: $< M\, y_1, Ky_2> \; = \; < y_1, M^* K\, y_2 >$

$$= \; <[y_1, y_2]\, , M^* K >$$

$$= \; < M[\, y_1, y_2], K > \quad .$$

5.4 Necessary Conditions for Parameter Optimality

Consider the system

$$y \; = \; y_o \; + \; W\, u \; . \qquad 5.4.1$$

We will assume that all the parameters that we wish to vary are expressed as gains from the output y to the input u. This is a restriction, but it covers a very large number of cases, as will be shown.

Then $$u = -K y \ . \qquad 5.4.2$$

The operator K (typically a matrix of gains though not necessarily) is a restricted operator, of the type discussed in the previous section. We make the further assumption that, within the restricted class, K is arbitrary, i.e. all the parameters are free to be varied arbitrarily. Then, for a given K, the closed-loop system becomes

$$y = y_o - W K y \qquad 5.4.3$$

or $$y = (I + W K)^{-1} y_o \ . \qquad 5.4.4$$

We wish to choose K to minimise

$$J = \tfrac{1}{2} \langle y, Q y \rangle + \tfrac{1}{2} \langle u, R u \rangle \qquad 5.4.5$$

$$= \tfrac{1}{2} \langle y, (Q + K^* R K) y \rangle \ . \qquad 5.4.6$$

Now J is definitely not quadratic in K. Hence, we shall employ a variational technique to determine stationary values of K, rather than try to find explicit minimum conditions. Second order effects are investigated in a later section.

For an arbitrary change in the control δu, the first order variation in cost δJ is given by

$$\delta J = \langle R u + W^* Q y, \delta u \rangle \qquad 5.4.7$$

(see Section 2.2). But from 5.4.2

$$\delta u = - K\,\delta y - \delta K\,y \,. \qquad 5.4.8$$

Hence
$$- \delta J = \langle R\,u + W^{*} Q\,y,\ \delta K\,y + K\,\delta y \rangle \qquad 5.4.9$$

$$= \langle - R\,K\,y + W^{*} Q\,y,\ \delta K\,y + K\,\delta y \rangle$$

$$= \langle - R\,K\,y,\ \delta K\,y \rangle + \langle - R\,K\,y,\ K\,\delta y \rangle$$

$$+ \langle W^{*} Q\,y,\ - \delta u \rangle \,. \qquad 5.4.10$$

However, from equation 5.4.1

$$\delta y = W\,\delta u \,. \qquad 5.4.11$$

$$\therefore \qquad - \delta J = \langle - R\,K\,y,\ \delta K\,y \rangle + \langle - R\,K\,y,\ K\,W\,\delta u \rangle$$

$$+ \langle W^{*} Q\,y,\ - \delta u \rangle$$

$$= \langle - R\,K\,y,\ \delta K\,y \rangle + \langle - R\,K\,y,\ K\,W\,\delta u \rangle$$

$$+ \langle Q\,y,\ - W\,\delta u \rangle$$

$$= \langle - R\,K\,y,\ \delta K\,y \rangle + \langle - (K^{*} R\,K + Q)y,\ W\,\delta u \rangle \,.$$
$$5.4.12$$

From 5.4.8, and 5.4.11,

$$\delta u = - K\,W\,\delta u - \delta K\,y \,.$$

i.e.
$$\delta u = - (I + K\,W)^{-1}\,\delta K\,y \,. \qquad 5.4.13$$

Hence
$$\delta J = \langle R\,K\,y,\ \delta K\,y \rangle + \langle (K^{*} R\,K + Q)y,\# W(I + KW)^{-1} \delta K y \rangle$$
$$5.4.14$$

$$= \langle [R\,K\,y,\ y],\ \delta K\rangle \ - \ \langle [F^*(Q + K^*\,R\,K)y,\ y],\ \delta K\rangle \qquad 5.4.15$$

where $$F = W(I + K\,W)^{-1} \qquad 5.4.16$$

and the [,] notation is consistent with Section 5.3. For a stationary value of K, we have

$$\delta J = 0 .$$

However, since δK is completely arbitrary, by assumption,

$$[R\,K\,y,\ y] = [F^*(Q + K^*\,R\,K)y,\ y] \ . \qquad 5.4.17$$

From the lemma 5.3.1

$$[R\,K\,y,\ y] = R\,K[y,\ y] \ , \qquad 5.4.18$$

if R K lies in the restricted operator class, which we will assume. X X

Hence $$R\,K[y,\ y] = [F^*(Q + K^*\,R\,K)y,\ y] \qquad 5.4.19$$

We shall denote the functional [y, y] by G, and by analogy with the case where [y, y] simply becomes a square matrix, we shall call it the Gramian. Furthermore, the invertability of G will be assumed. In the case $y \in R^n \otimes L_2$, this is allowable when the component functions $y_i(t)$ are independent.

For a gain K, the closed loop system is shown pictorially in Figure 5.4.1.

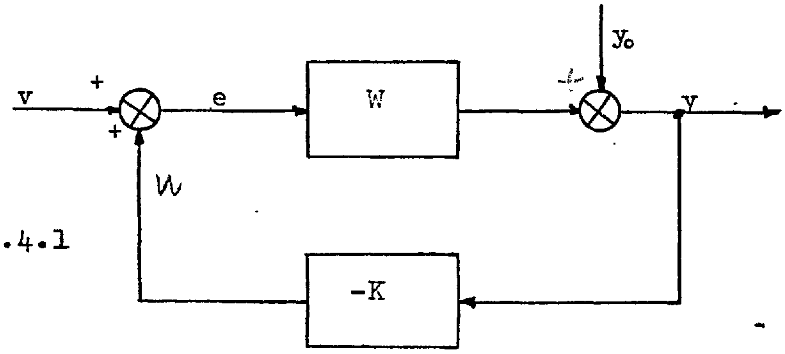


fig. 5.4.1

We proceed to derive the transfer operator from the subsidiary input v to the output y.

$$e = v + u$$

$$= v - K\,y$$

$$= v - K\,y_o - K\,W\,e \quad .$$

$$(I + K\,W)e = v - K\,y_o \quad .$$

$$e = (I + K\,W)^{-1}\,v - (I + K\,W)^{-1}\,K\,y_o \quad .$$

But $$y = y_o + W\,e \; .$$

$$y = W(I + K\,W)^{-1}\,v + y_o - (I + K\,W)^{-1}\,K\,y_o \quad .$$

Thus $$F = W(I + K\,W)^{-1}$$

is the closed loop transfer operator from v to y. Note that also

$$y = y_o + W(u + v)$$

$$= y_o - W\,K\,y + W\,v \quad .$$

i.e. $$y = (I + W K)^{-1} W v + (I + W K)^{-1} y_o$$

Hence $$F = W(I + K W)^{-1} = (I + W K)^{-1} W \qquad 5.4.20$$

5.5 Computational Algorithms I

The necessary condition of optimality, given by equation 5.4.19 is quite a complicated expression in K, and in general no explicit algebraic solution is possible. Iterative techniques are therefore desirable. Now 5.4.19 becomes

$$R K G = [F^*(Q + K^* R K)y, y] \qquad 5.5.1$$

where $$G = [y, y] . \qquad 5.5.2$$

Denote $$F^*(Q + K^* R K) y = \phi$$

$F = W(I+KW)^{-1}$
$= (I+KW)^{-1} W$

and $$\Psi = [\phi, y] .$$

Then $$R K G = \Psi . \qquad 5.5.3$$

From 5.4.15 $$\delta J = \langle R K G - \underline{\Psi}, \delta K \rangle , \qquad 5.5.4$$

$$= \langle g_K, \delta K \rangle .$$

The following algorithms are proposed.

Algorithm 5.5.1

(want M = Ψ - Q [illegible]

$$K_{i+1} = R^{-1} \underline{\Psi}_i \; G_i^{-1} \qquad 5.5.5$$

Algorithm 5.5.2

$$K_{i+1} = (1 - \varepsilon) K_i + \varepsilon R^{-1} \underline{\Psi}_i \; G_i^{-1} \qquad 5.5.6$$

$$0 < \varepsilon \leqq 1 \; .$$

Now $\quad R\,K\,G = \underline{\Psi} \; .$

$$\therefore \qquad (R + M) K\, G = M\, K\, G + \underline{\Psi}$$

Algorithm 5.5.3

$$K_{i+1} = (R + M)^{-1} (\underline{\Psi}_i \; G_i^{-1} + M\, K_i) \qquad 5.5.7$$

M positive semi-definite in general.

Algorithm 5.5.4

$$K_{i+1} = K_i (1 - \varepsilon) + \varepsilon (R + M)^{-1} [\underline{\Psi}_i \; G_i^{-1} + M\, K_i] \quad 5.5.8$$

Further variations are obtained by allowing M and ε to vary from iteration to iteration. These algorithms were first obtained by observation of the similarity of the basic equations with the state-space algorithms of Section 5.2. However, we now proceed to show the convergence properties of these algorithms, and methods for choosing ε and M.

We first digress to make the following two observations.

Lemma 5.5.1: If, for a certain disturbance y_o, K is optimal gain (in the above sense), then K is also optimal for any disturbance αy_o, where α is a scalar.

Proof: Since $y = (I + W K)^{-1} y_o$,

then αy_o will produce an output αy for the same K. But

$$\begin{aligned} g_k &= R K[\alpha y, \alpha y] - [F^*(Q + K^* R K)\alpha y, \alpha y] \\ &= \alpha^2 \{ R K[y, y] - [F^*(Q + K^* R K)y, y] \} \\ &= 0 , \end{aligned}$$

if K is optimal for the disturbance y_o.

This property justifies normalising the gradient with respect to the magnitude of the signals involved, corresponding to multiplying by G^{-1}. If G^{-1} is difficult to calculate, as in distributed systems, then it may still be useful to perform this normalisation approximately, perhaps by dividing by the norm of G. The following lemma also proves useful in the sequel.

Lemma 5.5.2: If $A = (a_{ij})$, and $B = (b_{ij})$ are both positive semi-definite and symmetric matrices of the same size, then $\operatorname{tr} A B \geq 0$.

Proof: By Schur's theorem [B 1; p.94]

$$C = (c_{ij}) = (a_{ij}\, b_{ij})$$

is positive semi-definite.

But $\operatorname{tr}(A\,B) = \sum_{i,j} a_{ij}\, b_{ji}$

$= \sum_{i,j} a_{ij}\, b_{ij}$ by symmetry

$= \sum_{i,j} c_{ij}$.

However $x^T C\, x \geqq 0 \qquad \forall\, x$.

Choose each $x_i = 1$.

Then $x^T C\, x = \sum_{i,j} c_{ij}$

$\geqq 0$.

i.e. $\operatorname{tr}(A\,B) \geqq 0$. 5.5.9

<u>Property 5.5.1</u>: The algorithm 5.5.2 produces, to first order, a decrease in cost at each iteration.

<u>Proof</u>: $K_{n+1} = K_n(1 - \varepsilon) + \varepsilon R^{-1} \Psi G^{-1}$

(Note that algorithm 5.5.1 is just the particular case of algorithm 5.5.2, when $\varepsilon = 1$).

$$K_{n+1} = K_n - \varepsilon(K_n - R^{-1}\Psi G^{-1})$$

$$= K_n - \varepsilon R^{-1}(R K_n G - \Psi)G^{-1}. \qquad 5.5.10$$

i.e. $$K = - \varepsilon R^{-1} g_k G^{-1} .$$

The first order change in cost is given by

$$\delta J = < g_k, K >$$

$$= -< g_k, \varepsilon R^{-1} g_k G^{-1} >$$

$$= - \varepsilon \, tr \, (g_k^{*} R^{-1} g_k) G^{-1} . \qquad 5.5.11$$

In the case of K a matrix of gains, both $g_k^T R^{-1} g_k$, and G^{-1} are positive semi-definite matrices. Hence by the lemma 5.5.2

$$tr \, (g_k^T R^{-1} g_k) . G^{-1} \geq 0$$

$$\therefore \qquad - \delta J \geqq 0 .$$

Since $\delta J = < g_k, K >$

then $\delta J = 0$ only at a stationary value of K.

i.e. $-\delta J > 0$ at each non-stationary iteration.

To ensure an actual decrease in cost at each iteration, ε should be chosen appropriately. To prove the theorem in general, we use the following argument.

$$\delta J = -\varepsilon < g_k, R^{-1} g_k G^{-1} >$$

$$= -\varepsilon < g_k G^{-1} G, R^{-1} g_k G^{-1} >$$

$$= -\varepsilon < G, G^{-1*} g_k^{*} R^{-1} g_k G^{-1} >$$

Now the operator $G^{-1*} g_k^{*} R^{-1} g_k G^{-1}$ is positive semi-definite.

Hence $$\delta J = -\varepsilon < [y, y], G^{-1*} g_k^{*} R^{-1} g_k G^{-1} >$$

$$= -\varepsilon < y, (G^{-1*} G_k^{*} R^{-1} g_k G^{-1}) y >$$

$$\leqq 0 .$$

Theorem 5.5.1: Consider the continuous system

$$\dot{x} = A x + B u \quad ; \qquad x(o) = x_o \qquad 5.5.12$$

where $x \in R^n$, $u \in R^m$, and A and B are constant matrices.

Let $$J = \int_o^\infty x^T Q x + u^T R u \, dt \;, \qquad 5.5.13$$

and $u = - K x$ (Q, R and K are constant matrices).

Then algorithm 5.5.1 is equivalent to algorithm 5.2.1.

Proof: $$K_n = - R^{-1} \Psi G^{-1}$$

and $$G = \int_o^\infty x(t) \, x^T(t) \, dt \; .$$

For any gain K, the closed loop response is given by

$$\dot{x} = (A - B K)x \quad ; \quad x(o) = x_o .$$

i.e. $x(t) = \Phi_k \, x_o$

where $\Phi_k = e^{(A - B K)t}$.

Hence $F_k(t) = \Phi_k(t) \, . \, B$

$$\Psi = \int_o^{\infty} \left(\int_t^{\infty} B^T \Phi_k^T(\tau - t)(Q + K^T RK) \Phi_k(\tau) \, x_o \, d\tau \right) x^T(t) \, dt$$

But $\Phi_k(\tau) = \Phi_k(\tau - t) \Phi_k(t)$.

Hence
$$\Psi = \int_o^{\infty} \left(\int_t^{\infty} B^T \Phi_k^T(\tau - t)(Q + K^T RK) \Phi_k(\tau - t) \Phi_k(t) x_o \, d\tau \right) x^T(t) \, dt$$

$$= \int_o^{\infty} B^T \left(\int_o^{\infty} \Phi_k^T(\tau - t) \, (Q + K^T RK) \Phi_k(\tau - t) d\tau \right) x(t) . x^T(t) \, dt .$$

Make a change of variable $\xi = \tau - t$. Then

$$\Psi = \int_o^{\infty} B^T \left(\int_o^{\infty} \Phi_k^T(\xi)(Q + K^T RK) \Phi_k(\xi) \, d\xi \right) x(t) \, x^T(t) \, dt .$$

But
$$P_k = \int_o^{\infty} \Phi_k^T(\xi)(Q + K^T RK) \Phi_k(\xi) \, d\xi$$

$\therefore$

$$\Psi = B^T \, P_k \int_o^{\infty} x(t) \, x^T(t) \, dt .$$

ψ regards [illegible] ?

[illegible]

But
$$\begin{aligned} K_{n+1} &= + R^{-1} \Psi_n \, G_n^{-1} \\ &= + R^{-1} \, B^T \, P_n \, G_n \, G_n^{-1} \\ &= + R^{-1} \, B^T \, P_n \end{aligned}$$

as for algorithm 5.2.1.

From the descent property 5.5.1, and theorem 5.5.1, the following properties may be expected.

1. If the outputs from which feedback is derived represent the dominant states, then the algorithm 5.5.1 may be expected to converge as well as algorithm 5.5.2. (Note that "dominance" is a property of the closed loop.)

2. If the output is not a very good approximation to the dominant states, then algorithm 5.5.2, with a small enough value of $\varepsilon > 0$, will provide a safer, but slower algorithm. It may be convenient to choose ε to minimise the cost in each search direction. However, this can involve rather a lot of cost calculations and trajectory evaluations. It is probably better to modify ε in a simpler way, or fix it "a priori" to some estimated safe value.

To illustrate these assertions, we shall present some simple examples. When k is a simple scalar gain, it is more convenient for hand computation to use the following expression which we first derive. For continuous deterministic problems, equation 5.5.3

$$R\,K\,G = \underline{\Psi}$$

becomes
$$r\,k \int_0^T y^2\,dt = \int_0^T \left(\int_t^T F(\tau - t)(q + rk^2)y(\tau)d\tau\right)y(t)dt$$

$$= (q + rk^2) \int_0^T y(t) \left(\int_0^t F(t - \tau)y(\tau)d\tau\right)dt\ .$$

But
$$y(s) = (1 + W(s)\ .\ k)^{-1}\,y_o(s)\ .$$

Hence
$$\frac{dy}{dk}(s) = -\,(1 + W(s)k)^{-2}\,W(s)\,y_o(s)$$

$$= -\,F(s)\,y(s)\ .$$

$$\therefore\quad r\,k \int_0^T y^2\,dt = -\,(q + rk^2)\int_0^T y(t)\,\frac{dy}{dk}(t)\,dt$$

$$= -\,\frac{(q + rk^2)}{2}\,\frac{d}{dk}\int_0^T y^2(t)\,dt\ .$$

Hence algorithm 5.5.1 becomes

$$K_{n+1} = -\left(\frac{q + rk_n^{\ 2}}{2r}\right)\frac{\frac{d}{dk}(G_n)}{G_n} \qquad 5.5.14$$

where
$$G_n = \int_0^T y_n^{\ 2}\,dt\ . \qquad 5.5.15$$

Example 5.5.1: 2 state example ; 1 dominant state.

$$y(s) = \frac{1}{s(1+s)} \; u(s) + \frac{1}{s(1+s)}$$

$$u(s) = -\,k\,y(s)$$

Then $$y(s) = \frac{1}{s^2 + s + k} \; .$$

We wish to choose the scalar k to minimise

$$J = \int_0^\infty q\,x^2 + u^2\,dt \; ,$$

with respect to k. Using Parseval's theorem, we obtain

$$G = \frac{1}{2k} \; .$$

Hence $$\frac{dG}{dk} = -\,\frac{1}{2k^2}$$

Equation 5.5.14 becomes

$$k_{n+1} = \frac{q + k^2}{2} \cdot \frac{\frac{1}{2k^2}}{2k}$$

$$= \frac{q}{2k_n} + \frac{k_n}{2} \; ,$$

exactly as for the simple single state case of example 5.2.1.

These algorithms may be useful for systems with non-rational transfer functions, as the next few examples show.

Example 5.5.2: $W(t) = \frac{1}{\sqrt{t}}$

This form of weighting function comes from an idealised model of an open-circuited semi-infinite transmission line. If an impulse of voltage is impressed at the input, then $W(t)$ is the current response,

and $$W(s) = \sqrt{\frac{\pi}{s}} \ .$$

[ZD 1; p.402.] We desire to design a gain linking the output (current signal) to the input (voltage signal) to regulate the current. The block diagram of the closed loop system is shown in Figure 5.5.1.

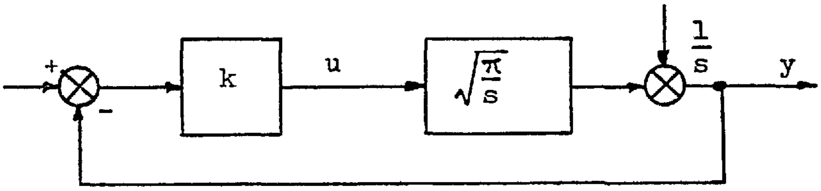


fig. 5.5.1

We assume a step disturbance $(y_o(s) = \frac{1}{s})$, and optimise

$$J = \int_0^T q\, y^2 + u^2 \, dt \ .$$

Now $$y(s) = (1 + \frac{k}{\sqrt{s}})^{-1} \frac{1}{s}$$

$$= \frac{1}{\sqrt{s}(\sqrt{s} + k)} \ .$$

From tables [CRC 1] ,

$$y(t) = e^{k^2 t} \operatorname{erfc} (k\sqrt{t})$$

$$\bar{y}(t) = e^{k^2 t} \operatorname{erfc} (k\sqrt{t})$$

where $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$

$$= \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-v^2} dv .$$

Hence, algorithm 5.5.12 becomes

$$k_{n+1} = - (\frac{1}{r} + k_n^2) \frac{\int_0^T y \frac{dy}{dk} dt}{\int_0^T y^2 dt} .$$

In this case,

$$\frac{dy}{dk} = \frac{d}{dk} (e^{k^2 t} \operatorname{erfc} (k\sqrt{t}))$$

$$= 2 k t y - 2\sqrt{\frac{t}{\pi}} .$$

For computation, a convenient numerical technique is to use the first few terms of the asymptotic expansion :

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \frac{e^{-x^2}}{2x} (1 - \frac{1}{2x^2} + \frac{1.3}{(2x^2)^2} - \frac{1.3.5}{(2x^2)^3} + \ldots)$$

Computational results are plotted in Figures 5.5.2-3. The computational difficulties with this example are caused by the infinite discontinuity of the weighting function, and the fact that

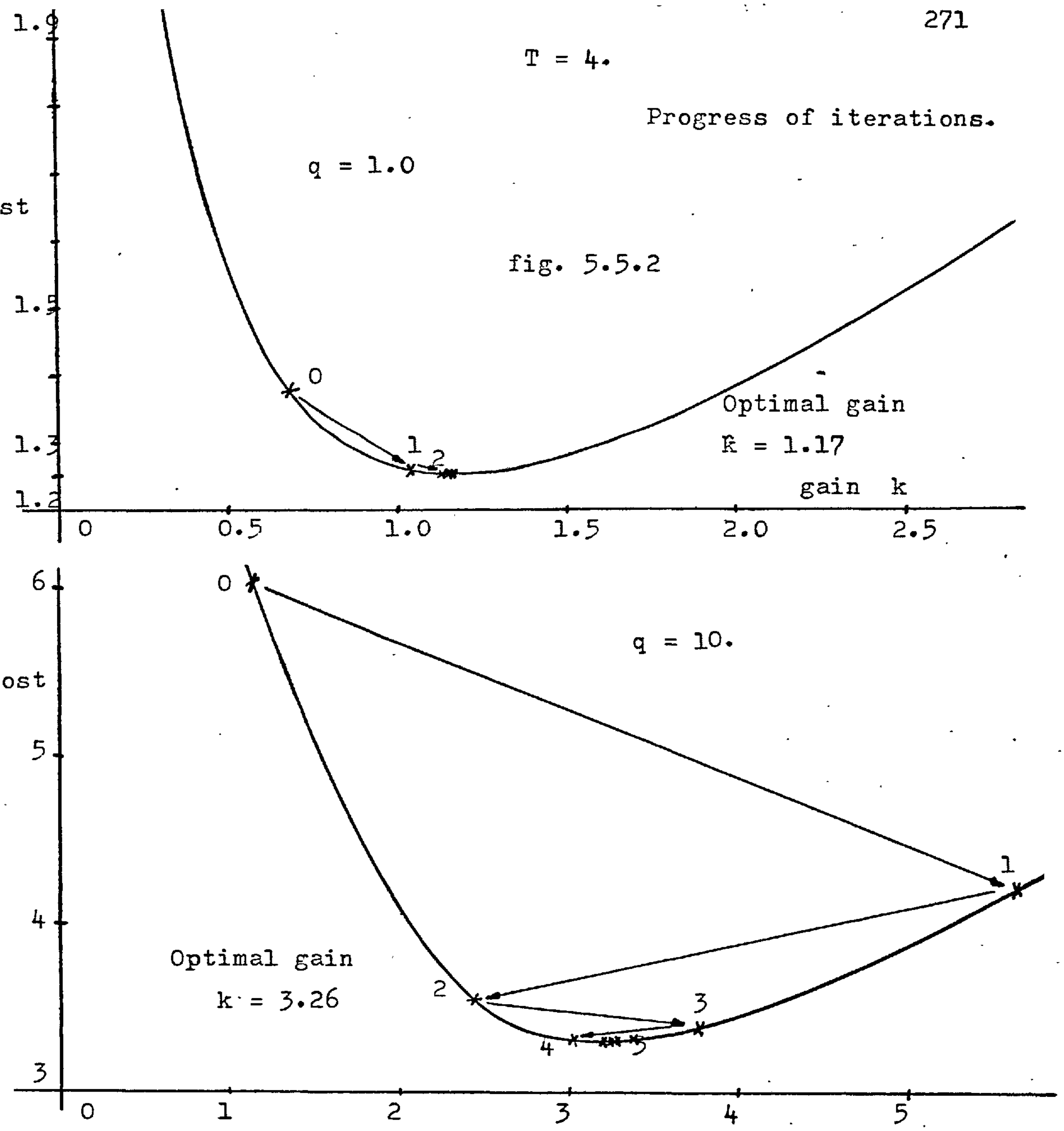


fig. 5.5.2

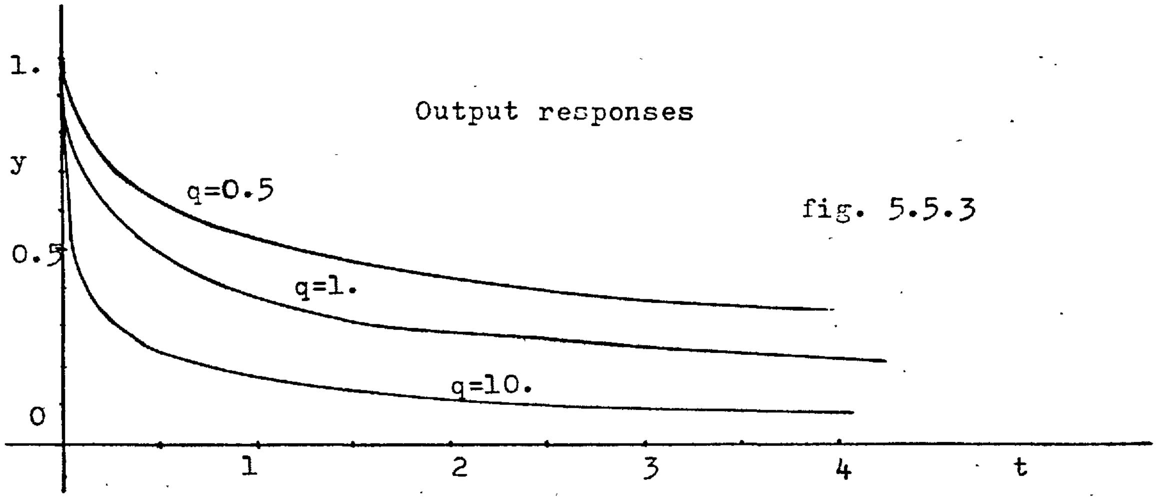


fig. 5.5.3

$$I = \int_{0}^{\infty} y^{2}\, dt$$

does not converge for any gain k. Bounded weighting functions enable standard programs to be written, which simply evaluate the quantities of algorithm 5.5.1 by numerical integration techniques. The following two examples illustrate this program.

Example 5.5.3

$$W(s) = \frac{e^{-4s}}{(s(1 + 0.5s))}$$

$$y_{o}(s) = \frac{1}{s} \quad .$$

We choose a scalar gain k to minimise

$$J = \int_{0}^{\infty} y^{2} + u^{2}\, dt$$

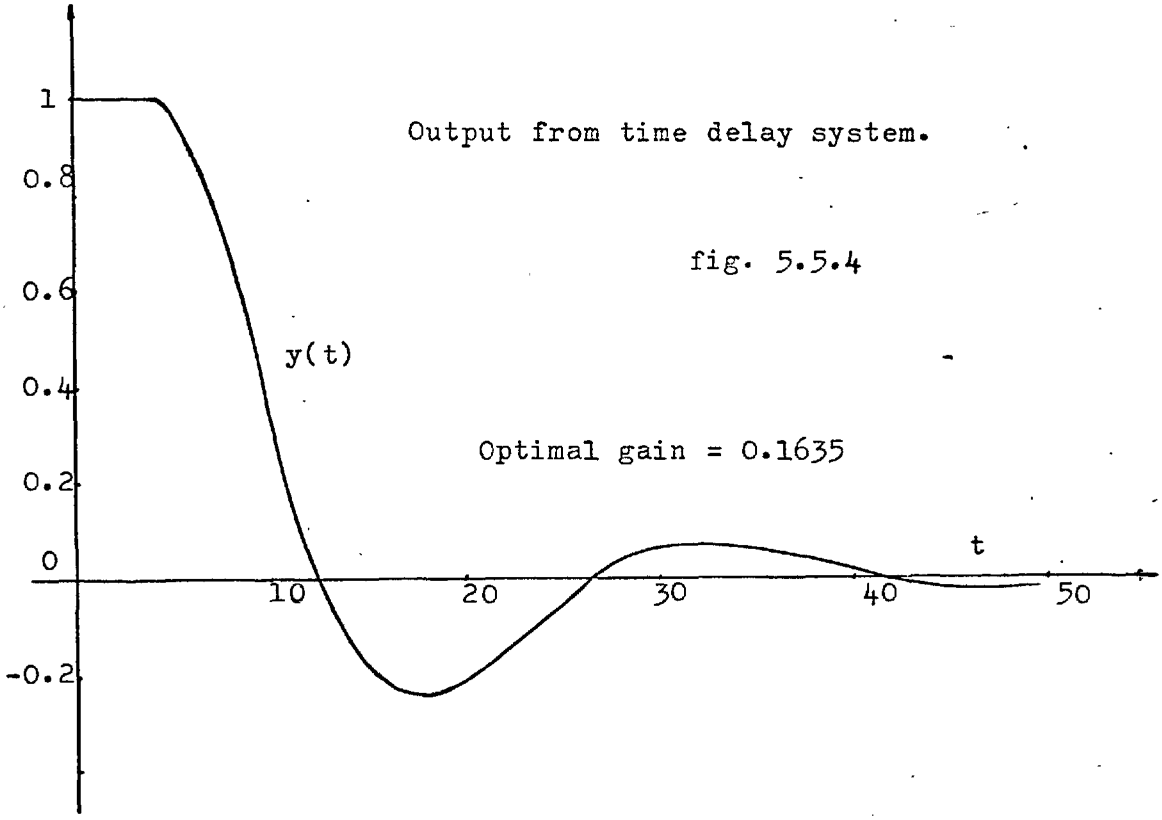
where $$y(s) = \frac{1 + 0.5s}{0.5s^{2} + s + ke^{-4s}} \quad .$$

The resulting output response is plotted in Figure 5.5.4.

Example 5.5.4

$$y_{o_1}(s) = y_{o_2}(s) = \frac{1}{s}$$

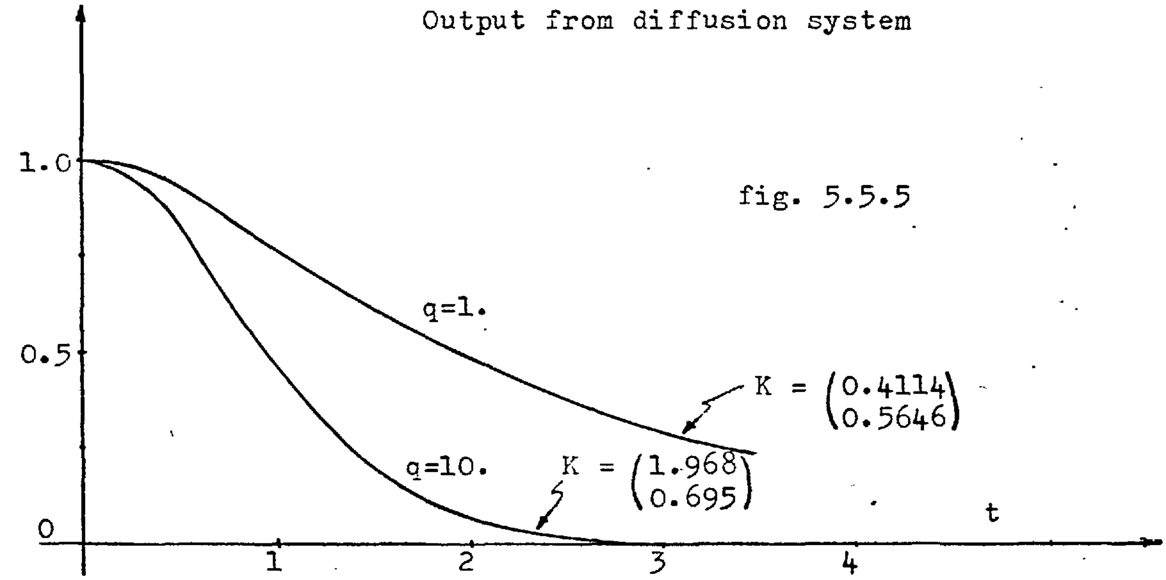
$$W(s) = \begin{pmatrix} \frac{e^{-\sqrt{s}}}{s} \\ \\ \frac{1}{s} \end{pmatrix}$$



Output from time delay system.

fig. 5.5.4

Optimal gain = 0.1635



Output from diffusion system

fig. 5.5.5

Design a gain vector

$$K = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} \quad , \text{ to minimise}$$

$$J = \int_0^\infty q\, y_1^{\,2} + u^2 \, dt$$

where
$$y_1(s) = \frac{1}{s} + \frac{e^{-\sqrt{s}}}{s} \; u(s)$$

$$y_2(s) = \frac{1}{s} + \frac{1}{s}\, u(s) \; .$$

This is the same system as for example 2.4.6. The initial disturbance represents a transition from one steady-state operating condition to another. For various values of q, the closed loop trajectories are calculated and plotted in Figure 5.5.5. These results are compared with the exact optimal trajectories calculated in Section 2.4, corresponding sets of trajectories are seen to be virtually identical. This demonstrates that for this system, an infinite dimensional state-space need not imply any sophisticated control law for good performance, though, of course, the "two gains" control law will not exhibit the infinite gain margin property.

5.6 Computational Algorithms II

So far, algorithms 5.5.1-2 have been discussed. The other two algorithms 5.5.3-4 prove very useful, especially for discrete time systems. Once again, algorithm 5.5.3 is a particular case of 5.5.4, for $\varepsilon = 1$, and for $M = 0$, we obtain algorithms 5.5.1-2.

Property 5.6.1: The algorithm 5.5.4 produces, to first order, a decrease in cost at each iteration, if M is positive semi-definite.

Proof:

$$\begin{aligned} K_{i+1} &= K_i - \varepsilon K_i - \varepsilon (R + M)^{-1} [\Psi G - MK_i] \\ &= K_i - \varepsilon (R + M)^{-1} [(R + M)K_i + \Psi G^{-1} - MK_i] \\ &= K_i - \varepsilon (R + M)^{-1} [RK_i + \Psi G^{-1}] \\ &= K_i - \varepsilon (R + M)^{-1} g_k G^{-1} . \qquad 5.6.1 \end{aligned}$$

$$\begin{aligned} \delta J &= < g_k, \delta K > \\ &= -\varepsilon < g_k, (R + M)^{-1} g_k G^{-1} > \\ &= -\varepsilon < y, (G^{-1*} g_k^* (R + M)^{-1} g_k G^{-1}) y > \\ &\leqq 0 \quad \text{as for the proof of property 5.5.1,} \end{aligned}$$

if M is positive semi-definite, since then $R + M$ is guaranteed positive definite if R is. In fact, all that is required is $R + M$ positive definite.

We now consider the effects of second order terms in the cost expression, to obtain an estimate of the convergence rate, and a method of choosing M. Now

$$J = < y, Q y > + < u, R u > .$$

From Section 2.2, we have

$$J + \Delta J = J + 2 < \Delta u, g > + < \Delta u, A \Delta u >$$

where $g = R u + W^* Q y$

and $A = R + W^* Q W$.

But $u = - K y$

$$u + \Delta u = - K y - \Delta K y - K \Delta y - \Delta K \Delta y$$

i.e. $\Delta u = - \Delta K y - K \Delta y - \Delta K \Delta y$.

Also $\Delta y = W \Delta u$

$\therefore$ $\Delta u = - \Delta K y - K W \Delta u - \Delta K W \Delta u$

$$(I + KW + \Delta K W) \Delta u = - \Delta K y \qquad 5.6.3$$

Put $D = I + KW$. 5.6.4

For a gain K, D is the return difference. Then

$$D \Delta u = - \Delta K y - \Delta K W \Delta u$$

$$\Delta u = - D^{-1}\Delta Ky - D^{-1}\Delta KW\Delta u$$

$$= - D^{-1}\Delta Ky - D^{-1}\Delta KW(-D^{-1}\Delta Ky - D^{-1}\Delta KW\Delta u)$$

$$= - D^{-1}\Delta Ky + D^{-1}\Delta KWD^{-1}\Delta Ky + \text{3rd order terms.} \qquad 5.6.5$$

Then $$\Delta J = 2\langle\Delta u, g\rangle + \langle\Delta u, A\Delta u\rangle$$

$$= - 2\langle D^{-1}\Delta Ky, g\rangle + 2\langle D^{-1}\Delta KWD^{-1}\Delta Ky, g\rangle$$

$$+ \langle D^{-1}\Delta Ky, AD^{-1}\Delta Ky\rangle + \text{3rd order terms,} \qquad 5.6.6$$

$$= - 2\langle \Delta Ky, D^{-1*}g\rangle + \langle 2\,\Delta KWD^{-1}\Delta Ky, D^{-1*}y\rangle$$

$$+ \langle \Delta Ky, D^{-1*}AD^{-1}\Delta Ky\rangle + \text{3rd order terms .} \qquad 5.6.7$$

We can now consider choosing ΔK, so that ΔJ is approximately minimised. This leads to Newton-Raphson type algorithms. Introduce a perturbation δΔK in ΔK, and consider the first order change in cost increment.

2.5.67

$$\delta\Delta J = - 2\langle \delta\Delta Ky, D^{-1*}g\rangle + 2\langle\delta\Delta KWD^{-1}\Delta Ky, D^{-1*}g\rangle$$

$$+ 2\langle\Delta KWD^{-1}\delta\Delta Ky, D^{-1*}g\rangle + 2\langle\delta\Delta Ky, D^{-1*}AD^{-1}\Delta Ky\rangle \quad 5.6.8$$

$$= - 2\langle\delta\Delta K, [D^{-1*}g, y]\rangle + 2\langle\delta\Delta K, [D^{-1*}g, WD^{-1}\Delta Ky]\rangle$$

$$+ 2\langle\delta\Delta K, [D^{-1*}W^{*}\Delta K^{*}D^{-1*}g, y]\rangle + 2\langle\delta\Delta K, [S\Delta Ky, y]\rangle$$

where $$S = D^{-1*} A D^{-1} \quad . \qquad 5.6.9$$

For an optimum, $\delta\Delta J = 0$ for arbitrary allowable $\delta\Delta K$. i.e.

$$[\, D^{-1*}g,\, y\,] = [\, D^{-1*}g,\, WD^{-1}\Delta Ky\,] + [\, D^{-1*}W^{*}\Delta K^{*}D^{-1*}g,\, y\,] + [S\Delta Ky,\, y] \,. \qquad 5.6.10$$

Now

$$D^{-1*}g = D^{-1*}(Ru + W^{*}Qy)$$

$$= D^{-1*}(-\, RKy + W^{*}Qy)$$

$$= D^{-1*}\, W^{*}Qy + D^{-1}W^{*}K^{*}RKy - D^{-1*}(W^{*}K^{*} + I)RKy \quad . \qquad 5.6.11$$

From equation 5.6.4,

$$D = I + KW \quad .$$

Hence

$$D^{*} = I + W^{*}K^{*} \,.$$

$$\therefore \qquad D^{-1*}g = D^{-1*}W^{*}(Q + K^{*}RK)y - RKy \quad .$$

Also

$$F = WD^{-1} \,.$$

$$\therefore \quad [\, D^{-1*}g, y\,] = [\, F^{*}(Q + K^{*}RK)y,\, y] - RKG$$

$$= \Psi - RKG \quad . \qquad 5.6.12$$

For a gain K, the Newton-Raphson step ΔK is given by

$$\Psi - RKG = [\, D^{-1*}g,\, WD^{-1}\Delta Ky\,] + [D^{-1*}W^{*}\Delta K^{*}D^{-1}g,\, y\,] + [\, S\Delta Ky,\, y\,] \,. \qquad 5.6.13$$

The self-adjoint operator S is given by

$$S = D^{-1*} A D^{-1}$$

$$= D^{-1*}(R + W^{*} Q W)D^{-1} .$$

As discussed in Chapter 2, the operator

$$A = R + W^{*} Q W$$

has a factorisation

$$A = T^{*} V T \qquad 5.6.14$$

where T is the optimal return difference, and V is an operator in the sub-space of instantaneous operators. If the gain K feeds back from all the state-variables of the system, and the appropriate stationarity assumptions are made, then

$$D = I + W K$$

is the optimal return difference (for arbitrary control variations), when K is optimal in its restricted class, i.e. at optimum

$$S^{o} = (D^{-1*} T^{*}) V (T D^{-1})$$

$$= V . \qquad 5.6.15$$

Also at optimum, $g = 0$. Hence, if the gain K is capable of achieving optimal control (for arbitrary control variations) then at the optimum, equation 5.6.13 becomes

$$\Psi - RKG = [V \Delta K y, y] \qquad 5.6.16$$

If V and ΔK are both restricted operators (e.g. time invariant) and commute with the [,] operation, then

$$- R K G = V \Delta K G \quad . \qquad 5.6.17$$

i.e.

$$\Delta K = V^{-1}(\Psi - R K G) G^{-1}$$

$$= V^{-1} \Psi G^{-1} - V^{-1} R K \quad . \qquad 5.6.18$$

i.e.

$$K_{n+1} = K_n + \Delta K$$

$$= K_n + V^{-1} \Psi G^{-1} - V^{-1} R K_n$$

$$= V^{-1}(\Psi G^{-1} + (V - R)K_n) \quad . \qquad 5.6.19$$

Near the global optimum, this algorithm is approximately Newton-Raphson.

For continuous time systems,

$$V = R \qquad 5.6.20$$

and so equation 5.6.19 reduces to

$$K_{n+1} = R^{-1} \Psi G^{-1} \qquad 5.6.21$$

- precisely algorithm 5.5.1.

For discrete time systems (state-space representation equation 2.8.2)

$$V = R + B^T P B \quad . \qquad 5.6.22$$

If, for a gain K, the closed loop transition operator is Φ , which maps from the space of initial conditions x_o into the trajectories x, then

$$\langle x_o, P\,x_o\rangle = \langle \Phi x_o, (Q + K^* R\,K)\,\Phi\,x_o\rangle . \qquad 5.6.23$$

$$= \langle x_o, (\Phi^*(Q + K^* R\,K)\Phi)\,x_o\rangle . \qquad 5.6.24$$

If $\qquad x_o = B\,u_o ,$

then $\qquad \langle x_o, P\,x_o\rangle = \langle u_o, B^T P\,B\,u_o\rangle$

$$= \langle u_o, (B^T\Phi^*(Q + K^* R\,K)\Phi B)\,u_o\rangle \qquad 5.6.25$$

But $\qquad \Phi B = F_t, \qquad 5.6.26$

the closed loop impulse response from input to output.

i.e. $\qquad B^T P\,B = F_t^*(Q + K^* R\,K)\,F_t . \qquad 5.6.27$

The subscript t indicates that F_t is a map from the space of vectors u_o into trajectories y, and not the operator F mapping input trajectories into output trajectories. However, if we permit the δ function as a valid signal

$$F_t\,u_o = F(u_o\,\delta) .$$

In this way, the vector u_o is formally extended onto a space of time functions, and the symbol $u_o\,\delta$ corresponds to a tensor multiplication $u_o \otimes \delta$.

We have now proved the following:

Theorem 5.6.1: The algorithms 5.2.1 and 5.2.2 exhibit second order (Newton-Raphson) convergence near the optimum.

To implement an exact Newton-Raphson algorithm for these "optimal gains" problems is difficult, since the hessian associated with K in equation 5.6.13 is very complicated. However, we shall be guided by the discrete time algorithm 5.2.2, and let

$$C = F_t^{*}(Q + K^{*} R K) F_t \qquad 5.6.27$$

and

$$M = \alpha C \quad . \qquad 5.6.28$$

Then, we may apply algorithm 5.5.3

$$K_{n+1} = (R + M_n)^{-1} \Psi G^{-1} + M K_n \quad . \qquad 5.6.29$$

For systems with complete state feedback, we set $\alpha = 1$ for discrete time while for continuous time systems $\alpha = 0$. With incomplete state feedback, α can be chosen to give a decrease in cost at each iteration. That this can be done is indicated by the following:

Property 5.6.1: For the algorithm implied by equations 5.6.27-9, a first order decrease in cost is obtained at each iteration.

Proof:

$$K_{n+1} = (R + \alpha C)^{-1}(\bar{\Psi} G^{-1} + \alpha C K_n)$$

$$= K_n - \frac{1}{\alpha}\left(\frac{1}{\alpha} R + C\right)^{-1} g_k G^{-1}$$

Since C is positive semi-definite, then if α is chosen positive, the result follows from property 5.5.2. The factor $\frac{1}{\alpha}$ takes the place of ε, and α should be chosen large enough to ensure that the first order term predominates.

We illustrate this algorithm with the following example:

Example 5.6.1: The discrete time system represented by

$$y(z) = y_o(z) + \left(\frac{a + bz}{z^2 + cz + d}\right) u(z)$$

was simulated on a PDP-9 computer, by means of the implied recursion relation. If we set

$$u(z) = - k\, y(z)$$

then $$y_j = (- c - bk)y_{j-1} + (- d - ak)y_{j-2} + y_{o_j} + cy_{o_{j-1}} + dy_{o_{j-2}}.$$

We wish to minimise

$$J = \sum_{j=0}^{N} q\, y_j^2 + u_j^2 \qquad (N = 50)$$

with respect to k, where N is an integer chosen large enough for the closed loop trajectories to have converged to zero. By recursion relationships similar to the above, the impulse response f, and the sensitivity F_y, can be calculated. With $\alpha = 1$, the algorithm of equation 5.6.29 becomes

$$K_{n+1} = \frac{(q + K_n^2)\,\frac{\langle y,\ F\,y\rangle}{\langle y,\ y\rangle} + c\,K_n}{1 + c}$$

where $$c = (q + K_n^2)\,\langle f,\ f\rangle \;.$$

Now we let $y_o(z) = \frac{1}{z - 1}$. By means of the conversational-mode facility, various constants were entered into the problem. The following particular cases are considered.

1. $q = 1,\ \ a = 0,\ \ b = 1,\ \ c = -1,\ \ d = 0,\ \ N = 50.$

Then $$W = \frac{1}{z - 1} \;.$$

For this case, we have a single state system, and the disturbance represents an initial condition on the plant. Hence the Riccati equation can be used to solve for the optimal gain:

$$p = p + 1 - \frac{p^2}{1 + p} \;.$$

$$p^2 - p - 1 = 0 \;.$$

$$p = \tfrac{1}{2} + \sqrt{\tfrac{5}{4}} \qquad \text{(taking the positive root)}$$

$$= 1.61803 \;.$$

$$k = \frac{p}{1 + p}$$

$$= 0.618034 \;.$$

Using our algorithm, the results of Table 5.6.1 were obtained, using different values of gain as a starting point.

Table 5.5.1

```
                 COST                  GAIN          LOOPS           BPB

              Ø.200000E+01        Ø.100000E+01         50       Ø.200000E+01
              Ø.162500E+01        Ø.666667E+00         50       Ø.162500E+01
              Ø.161804E+01        Ø.619048E+00         50       Ø.161804E+01
              Ø.161803E+01        Ø.618034E+00         50       Ø.161803E+01
              Ø.161803E+01        Ø.618034E+00         50       Ø.161803E+01
              Ø.161803E+01        Ø.618034E+00         50       Ø.161803E+01
↑S
NEXT    *       *

GK        1.8
NEXT    *       *
SIM
                 COST                  GAIN          LOOPS           BPB

              Ø.117778E+02        Ø.180000E+01         50       Ø.117778E+02
              Ø.186100E+01        Ø.921738E+00         50       Ø.186100E+01
              Ø.162117E+01        Ø.650472E+00         50       Ø.162117E+01
              Ø.161803E+01        Ø.618491E+00         50       Ø.161803E+01
              Ø.161803E+01        Ø.618034E+00         50       Ø.161803E+01
              Ø.161803E+01        Ø.618034E+00         50       Ø.161803E+01
  R↑S
NEXT    *       *

 GK        Ø.
 NEXT    *       *
 SIM
                 COST                  GAIN          LOOPS           BPB

               Ø.510000E+02        Ø.000000E+00         50       Ø.300000E+02
               Ø.171459E+01        Ø.806452E+00         50       Ø.171459E+01
               Ø.161859E+01        Ø.631621E+00         50       Ø.161859E+01
               Ø.161803E+01        Ø.618116E+00         50       Ø.161803E+01
               Ø.161803E+01        Ø.618034E+00         50       Ø.161803E+01
               Ø.161803E+01        Ø.618034E+00         50       Ø.161803E+01
```

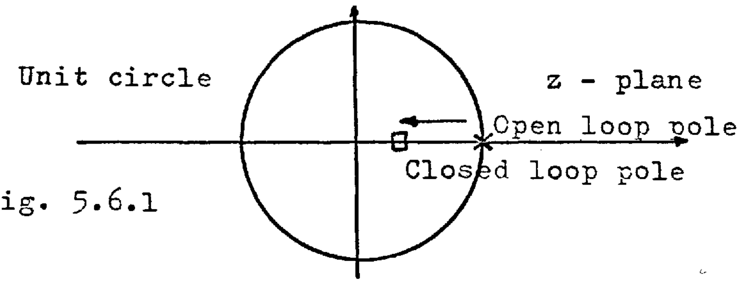


fig. 5.6.1

This procedure can be viewed as the shifting of a pole in the z-plane as shown in fig. 5.6.1.

As expected, for this example, the algorithm has excellent convergence properties, as it is identical to algorithm 5.2.2, which is second-order near the optimum.

If the parameters a and d are now given non-zero values, then a gain k from output to input is no longer an optimal control law.

2. $q = 1$, $a = 0.1$, $c = -1.1$, $d = 0.1$, $b = 1.0$.

On the z-plane, we have introduced an extra pole and zero, as shown on Figure 5.6.2.

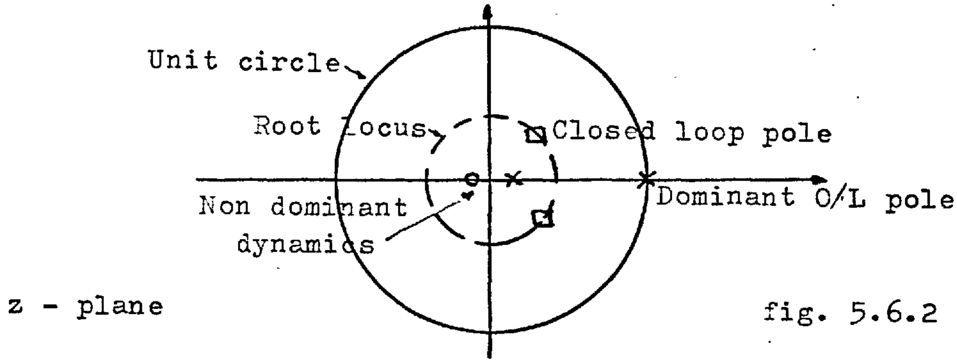


fig. 5.6.2

The iterations of the algorithm are shown in Table 5.6.2. The second order convergence property no longer holds, but good convergence is still obtained.

3. $q = 1$, $a = 0.2$, $c = -1.2$, $d = 0.2$, $b = 1.0$.

This example is similar to the last, but diverges even further from the single state case, in placing the extra pole and zero at +0.2 and -0.2 respectively. Iterations of our algorithm are shown in table 5.6.3.

```
NEXT    *         *
```

Table 5.6.2

```
A          0.1
NEXT    *         *
C         -1.1
NEXT    *         *
D          .1
NEXT    *         *
GK         1.2
NEXT    *         *
SIM
           COST                  GAIN           LOOPS          BPB

        0.258145E+01        0.120000E+01         50       0.256494E+01
        0.181366E+01        0.801648E+00         50       0.192350E+01
        0.171847E+01        0.648805E+00         50       0.186878E+01
        0.171400E+01        0.615520E+00         50       0.187412E+01
        0.171391E+01        0.610656E+00         50       0.187551E+01
        0.171390E+01        0.610024E+00         50       0.187571E+01
        0.171390E+01        0.609943E+00         50       0.187573E+01
        0.171390E+01        0.609933E+00         50       0.187573E+01
        0.171390E+01        0.609932E+00         50       0.187573E+01
        0.171390E+01        0.609932E+00         50       0.187573E+01
↑S
NEXT    *         *
```

Table 5.6.3

```
A          .2
NEXT    *         *
C          -1.2
NEXT    *         *
D          .2
NEXT    *         *
GK         0.
NEXT    *         *
SIM
           COST                  GAIN           LOOPS          BPB

        0.789713E+02        0.000000E+00         50       0.658854E+02
        0.191137E+01        0.551058E+00         50       0.236650E+01
        0.190831E+01        0.573976E+00         50       0.234809E+01
        0.190819E+01        0.578454E+00         50       0.234510E+01
        0.190819E+01        0.579383E+00         50       0.234451E+01
        0.190819E+01        0.579578E+00         50       0.234438E+01
        0.190819E+01        0.579619E+00         50       0.234436E+01
        0.190819E+01        0.579628E+00         50       0.234435E+01
        0.190819E+01        0.579629E+00         50       0.234435E+01
        0.190819E+01        0.579630E+00         50       0.234435E+01
        0.190819E+01        0.579630E+00         50       0.234435E+01
 ↑S
```

The algorithm 5.5.3, using a "slugging" matrix M, can also be usefully used for the design of continuous systems. For example, if the cost is chosen to be

$$J = \int_{0}^{\infty} y^T Q y \, dt \quad .$$

and the operator K and the dynamics permit a useful solution, then algorithm 5.5.3 is directly applicable, although algorithms 5.5.1-2 cannot be used. If the system were sampled, and all the states fed back, then the optimisation problem becomes valid, even though there is no cost on control. The following intuitive method is proposed. Let α be the estimated time response of the closed-loop system. (e.g. dominant time constant). Then set

$$\begin{aligned} M &= \alpha C \\ &= \alpha \left(F_t^{*} (Q + K^{*} R K) F_t\right) \\ &= \alpha F_t^{*} Q F_t , \end{aligned}$$

and use algorithm 5.5.3. Due to property 5.6.1, the larger α, the more likely the algorithm is to converge, but if α is too large, only slow convergence is obtained. The following simple example shows the application of the algorithm 5.5.3 on a continuous plant.

<u>Example 5.6.2</u>: $\quad y(s) = \frac{1}{s} + \frac{1}{s} . u(s)$

$$u(s) = - k \, y(s) .$$

$$J = \int_0^\infty y^2 + u^2 \, dt \quad .$$

This is the same system as in example 5.2.1. However, we now derive the discrete-time algorithm.

$$y(s) = f(s) = \frac{1}{s + k} \, .$$

$$\therefore \qquad < f, f > (1 + k^2) = \frac{1}{2k} + \frac{k}{2} \quad .$$

The algorithm becomes

$$K_{n+1} = \frac{\dfrac{1 + k_n^2}{2k_n} + \alpha \left(\dfrac{1 + k_n^2}{2k_n}\right) k_n}{1 + \alpha \left(\dfrac{1 + k_n^2}{2k_n}\right)} \quad .$$

For $\alpha = 0$, this reduces to the original continuous time algorithm.

For $\alpha = 1$, $$K_{n+1} = \frac{2(1 + k_n{}^2)}{2k_n + 1 + k_n{}^2} \quad .$$

From the direct solution the optimal k is 1. However, starting from $k_o = 0$, the iterations proceed as

$$k_1 = 2$$

$$k_2 = 1.111$$

$$k_3 = 1.00276 \; .$$

5.7 Compensator Design

This final section of the chapter suggests methods of closed loop system design using compensators, based on the algorithms presented above. The ideas are fairly intuitive, and much future work remains to be done in this field.

If the "dominant states" of a system are not available for direct feedback, but can be observed through some measurements, or if the achievable closed loop response obtained by gain feedback on the measured variables is not satisfactory, then a compensating filter is necessary.

The basic method of design is to add further dynamics to the given plant, so as to increase the number of independent inputs and outputs available for control purposes. Then, using this augmented plant, a quadratic cost function is formulated, and the above algorithms for computing optimal gains are used. The choice of added compensator dynamics is an open question, and the more parameters we wish to leave variable, the more flexible the design can be.

Some basic configurations that one can propose are shown in Figures 5.7.1-3.

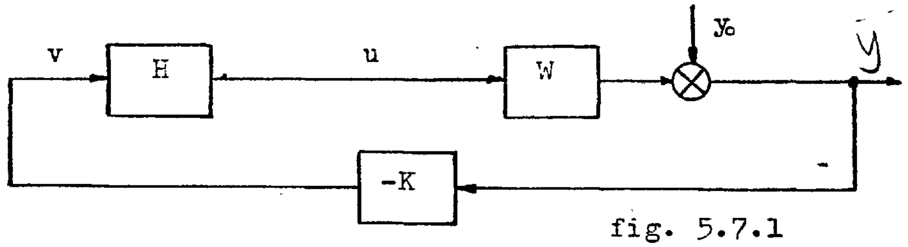


fig. 5.7.1

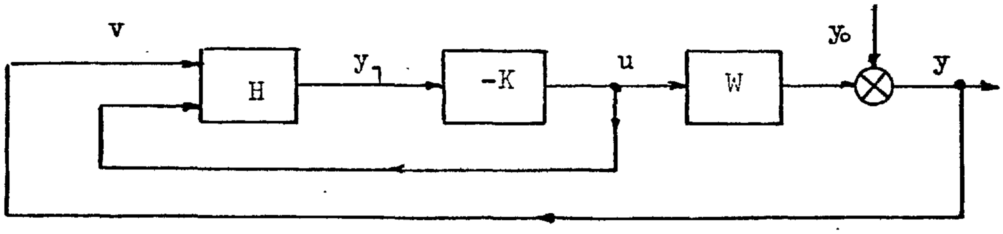


fig. 5.7.2

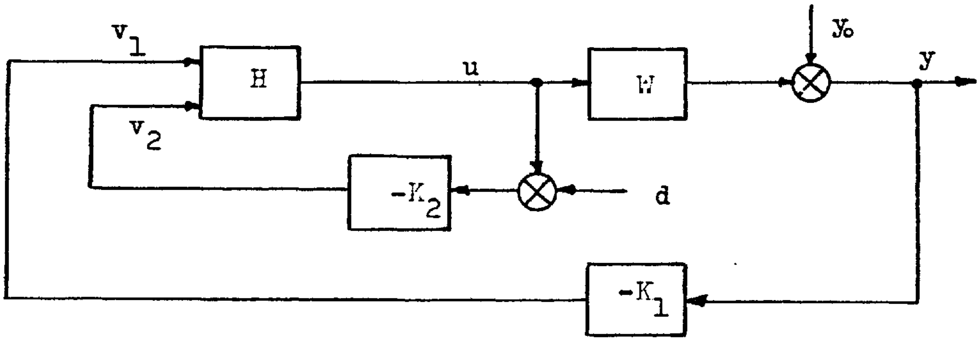


fig. 5.7.3

We shall concentrate the discussion on the case where u and y are single continuous time functions, although the extensions to sampled data and multivariable systems may be deduced in some cases.

The configuration of Figure 5.7.1 is particularly useful when the form of the compensator can be obtained from engineering experience. For instance, H can be designed to approximately cancel those states of W not available directly from the output. A particularly useful form of H is the three term controller.

i.e. $$H(s) = K_p + \frac{K_d\, s}{s + \alpha} + \frac{K_I}{s} \quad . \qquad 5.7.1$$

The coefficient α is either chosen " a priori ", to roll off high frequency noise, or else given as a fundamental limitation of the compensator.

$$H(s) = K_p + \frac{K_d(s + \alpha - \alpha)}{s + \alpha} + \frac{K_I}{s}$$

$$= (K_p + K_d) - \frac{\alpha K_d}{s + \alpha} + \frac{K_I}{s}$$

$$= K_1 + K_2 \left(\frac{\alpha}{s + \alpha}\right) + \frac{K_3}{s} \quad .$$

The closed loop system becomes:

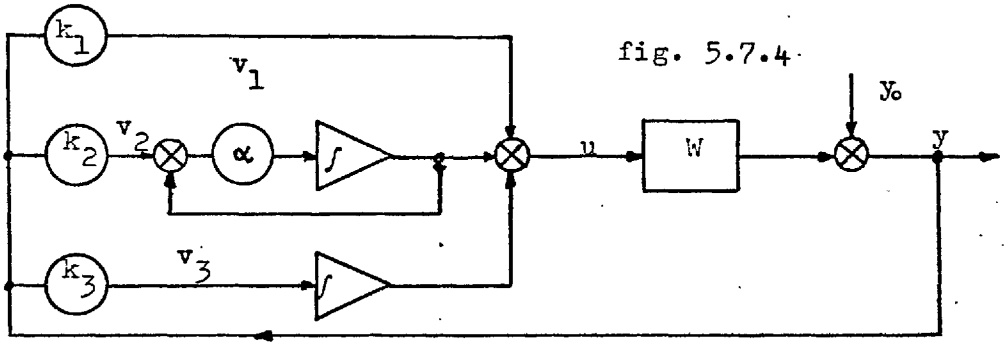


fig. 5.7.4

The cost is then formulated in terms of $\underset{\sim}{v}$ and y.

The structure of the form of Figure 5.7.2 enables the use of a simplified reduced order state space model of the plant. Using this model, one can design a Kalman filter, or Luenberger observer to estimate these states, to obtain the structure of Figure 5.7.5.

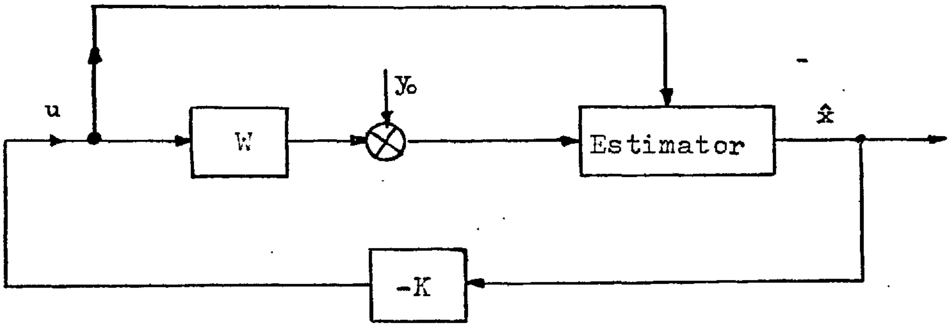


fig. 5.7.5

Now, costing x and u, an optimal K matrix can be calculated using the algorithms of the previous sections. Note that there is no stability problem with the Kalman filter, if only an approximate plant model is used, as distinct from feedback control design.

The configuration of Figure 5.7.3 enables complete generality to be achieved. In fact, if we use integrators as basic building blocks, then feedback and feedforward can generate any desired rational function of s. For the single-input/single-output case, we have a canonical structure made up of a chain of integrators as shown in Figure 5.7.6.

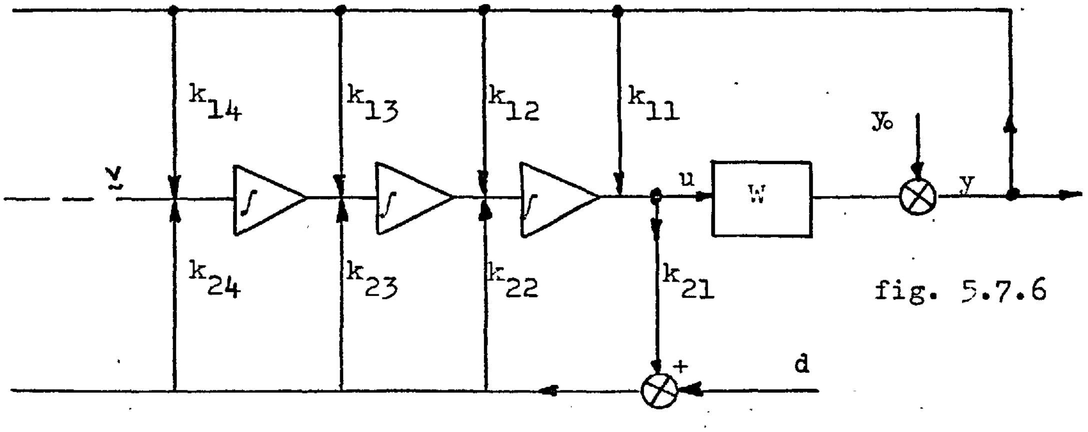


fig. 5.7.6

The chain of integrators can be terminated at any desired point. The disturbance d is added in to effectively prevent the gains k_{2i} from tending to infinity. The cost may be formulated in terms of $\bar{y}$ and u. Alternatively, the cost may be formulated in terms of the controls v and y, and then d need not be introduced. Further generalisations for multivariable systems would need to use canonical structures for multivariable compensators, which still form a research topic.

To sum up the whole chapter, it seems that the algorithms presented have the following advantages.

1. Whenever the optimisation problem approximates the all-state feedback infinite time problem in any sense, then convergence may be expected. Certainly convergence of the algorithm can be used as a test for the validity of an approximate state space model.

2. The algorithms can be implemented on an analog computer, using only integrators and multiplication. Small matrices must be handled, but no storage of trajectories is required, although a sensitivity model must be used. Alternatively, if trajectories can be stored, then the system model can be used as a sensitivity model.

Further research is needed to investigate better convergence criteria, the effect of varying the initial conditions, and the effect of non-linearities.

SUMMARY

The weighting function or system transfer approach is not new, but has fallen out of favour during the past decade. The purpose of this thesis has been to investigate how a weighting function approach can be used in the analysis and design of linear control systems. In many practical cases the use of a weighting function has computational advantages over the state space approach, especially when the dimension of the state space is large. Many theoretical advantages have also become evident.

The thesis has surveyed the field of the optimal control of linear systems with quadratic costs, and the dual problem of filtering. From this the subject of modelling errors, and restricted structure control has been investigated from a weighting function point of view. Simple problems, computed either by hand or machine, have been included as examples of the algorithms that have been proposed.

CONCLUSIONS AND FUTURE RESEARCH

System linearity is the main restrictive assumption used in this thesis. A secondary assumption is that of a quadratic performance index. However, under these restrictions a fairly unified theory has been developed. While there still remain many unresolved problems in this field, it is felt that useful computational tools have been developed for linear systems design.

A major conclusion of this thesis is that simple control structures, based on dominant dynamics, can produce near optimal performance. However, the concept of dominance depends on how the closed loop system is expected to perform. Further attention should be devoted to linear systems whose transfer functions are non-rational functions of s. These systems arise in distributed parameter problems, and appear to have some interesting properties. In particular, systems with structural resonances, such as the one possessing the square wave impulse response of section 2.7., present challenging control problems.

Most realistic models of engineering systems are non-linear, and a non-quadratic performance criterion is often specified. One common procedure for tackling such systems is to approximate the system (locally) by a linear system, and approximate the cost function (locally) by a quadratic function. The term 'locally' implies that the approximation is about a particular operating point. Hence future research could be directed at applying some of the algorithms presented in this thesis to non-linear systems. The descent algorithms of chapter 2 have already been applied in this way. However, the author feels that, with some re-derivation, the contraction algorithms of chapter 2 and the algorithms of chapter 5 could also be applied to non-linear systems.

A computational difficulty arises in the simulation of non-linear systems. If a model is specified in terms of a small number of ordinary differential equations or recurrence relations, then this is particularly easy to simulate. However the generalisation of the weighting function approach to non-linear systems by means of a Volterra functional series is extremely cumbersome computationally, since so much storage is required. A better method, though perhaps less general, is to model a system using only memoryless non-linearities and linear dynamics. The general representation of non-linear systems is of fundamental importance to the problem of identification, but no significant simplifications have yet been proposed.

BIBLIOGRAPHY

[A 1] Apostol,T. : 'Mathematical Analysis' ; Addison-Wesley 1957.

[A 2] Astrom,K.J. : 'Control Problems in Paper Making' ; I.B.M.Scientific omputing Symposium, Yorktown Heights, N.Y.,October,1964.

[A 3] Allwright,J. : Ph.D Thesis ; London, 1969.

[AG 1] Aizermann and Gantmacher : 'Absolute Stability of Regulator Systems' ; Holden-Day Inc. 1964.

[AR 1] Antosiewicz,H.A. and Rheinboldt,W.C. : 'Numerical Analysis and Functional Analysis' ; ch. 14 of 'Survey of Numerical Methods' ed. J.Todd;,p.485-517; McGraw-Hill, 1962.

[B 1] Bellman,R. : 'Introduction to Matrix Analysis' ; McGraw-Hill, 1960.

[BC 1] Bellman,R and Cooke,K.L. : 'Differential-Difference Equations' ; Academic Press. 1963.

[BM 1] Birkhoff,G. and MacLane,S. : 'A Survey of Modern Algebra' ; 3rd edition, Macmillan, 1965.

[C 1] Collatz,L. : 'Functional Analysis and Numerical Mathematics' ; Academic Press, 1966.

[C 2] Choksy,L. : 'Time Lag Systems' ; p.19-38 of 'Progress in Control Engineering Volume I', ed. Macmillan, Higgins, and Naslin. Heywood and Co.

[CRC1] CRC Standard Mathematical Tables 14th ed. Edited by S.M.Selby and B.Girling. The Chemical Rubber Company, 1965.

[D 1] Davis,M.C. : 'Factorising the Spectral Matrix' ; I.E.E.E. trans. on Auto. Control. Vol. AC-8 1963 ; p. 296-305.

[D 2] Dewey : 'Lecture Notes' .

[DS 1] Dunford,N. and Schwartz,J.T. : 'Linear Operators' Volume I ; Interscience.

[DS 2] Dunford,N. and Schwartz,J.T. : 'Linear Operators Volume II ' ; Interscience.

[FR 1] Fletcher,R. and Reeves,C.M. : 'Functional Minimisation by Conjugate Gradients' ; Computer Journal Vol.7 1964 ; p. 149-154.

[G 1] Guillemin,E.A. : 'Theory of Linear Physical Systems' ; Wiley, 1963.

[H 1] Halmos,P. : 'Finite-Dimensional Vector Spaces' ; 2nd. edition, Van Nostrand.

[H 2] Horvâth,J. : 'Topological Vector Spaces and Distributions' Vol. I. ; Addison-Wesley.

[H 3] Hayes,R.M. : 'Iterative Methods of Solving Linear Problems on Hilbert Spaces' ; N.B.S. Research Report. Applied Mathematics, series 39, 1954, p.71-104.

[H 4] Hsieh,H.C. : 'Synthesis of Adaptive Control Systems By Function Space Methods' ; p. 117-208 of 'Advances in Control Systems Vol.2' ; ed. C.T.Leondes. Academic Press, 1965.

[H 5] Horowitz,I.M. : 'Synthesis of Feedback Systems' ; Academic Press, 1963.

[K 1] Korevaar,J. : 'Mathematical Methods Vol. I. Linear Algebra/Normed Spaces/Distributions/Integration.'

[K 2] Kalman,R.E. : 'When is a Linear Control System Optimal?'; Trans.A.S.M.E. Vol.86, part D. March 1964. p. 1-10 .

[KB 1] Kalman,R.E. and Bucy,R.S. ; 'New Results in Linear Filtering and Prediction Theory' ; Trans. A.S.M.E. series D, J.Basic.Engrg. March 1961.. p. 95-108.

[KL 1] Kleinman,D.L. ; 'On an iterative Technique for Riccati Equation Computations' ; I.E.E.E. trans. on Auto. Control. Vol AC-13, p.114-115; Feb. 1968.

[L 1] Luenberger,D.G. : 'Observing the State of a Linear System' ; I.E.E.E. trans. on Military Electronics. Vol. MIL-8,p.74-80, April 1964.

[L 2] Luenberger,D.G. : 'Observers for Multivariable Systems' ; I.E.E.E. trans. on Auto. Control. Vol. AC-11, No.2 p. 190-197, April 1966.

[L 3] Luenberger,D.G. : 'A New Derivation of the Quadratic Loss Function Equation' ; I.E.E.E. trans. on Auto. Control. Vol. AC-10: p.202-3. April, 1965.

[L 4] Loève,M. : 'Probability Theory' ; 3rd. edn. Van Nostrand.

[LMW1] Lasdon,L.S., Mitter,S., and Waren,A.D. : 'The Method of Conjugate Gradients for Optimal Control Problems' ; I.E.E.E. trans. on Auto. Control, Vol. AC-11. June 1966. p. 904-5.

[M 1] Mikusinski,J. : 'Operational Calculus' ; Pergamon.

[M 2] Mayne, D.Q. : Lecture Notes.

[P 1] Paige, C.C. : 'Notes on Matrix Computations' ; University of London, Institute of Computer Science.

[P 2] Pagurek,B. : 'Sensitivity of the performance of Optimal Control Systems to Plant Parameter Variations' ; I.E.E.E. trans. on Auto. Control. Vol. AC-10. April, 1965. p. 178-180.

[RN 1] Riesz , and Sz-Nagy : 'Functional Analysis' ; Frederic Ungar Publishing Co. 1965 ed.

[SL 1] Sinnott,J.F., and Luenberger,D.G. : 'Solution of Optimal Control Problems by the Method of Conjugate Gradients' ; JACC. 1967 , p. 566-573.

[T 1] Taylor,A.E. : 'Introduction to Functional Analysis' ; Wiley.

[T 2] Truxal,J.G. : 'Identification of Process Dynamics' ; Ch. 3. of 'Adaptive Control Systems'. Ed. Mishkin and Braun. McGraw-Hill, 1961.

[V 1] Vulikh,B.Z. : 'Introduction to Functional Analysis for Scientists and Technologists' ; Pergamon, 1963.

[V 2] Varga,R.S. : 'Matrix Iterative Analysis' ; Prentice-Hall, 1962.

[W 1] Wilkinson,J.H. : 'The Algebraic Eigenvalue Problem' ; Oxford, 1965.

[W 2] Wiener,N. : 'The Extrapolation, Interpolation and Smoothing of Stationary Time Series' ; Wiley, 1958.

[Y 1] Youla, D.C. : 'On the Factorisation of Rational Matrices' ; I,E.E.E. trans. on Info. Theory; Vol. IT-7, 1961 ; p. 172-180.

[ZD 1] Zadeh,L.A. and Desoer,C.A. : 'Linear System Theory. The State Space Approach' ; McGraw-Hill, 1963.