

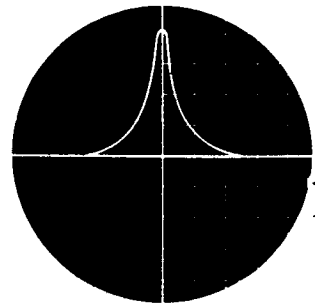
SEMI-ANNUAL STATUS REPORT  
VOLUME I

RESEARCH IN AND APPLICATION OF MODERN AUTOMATIC CONTROL THEORY  
TO NUCLEAR ROCKET DYNAMICS AND CONTROL

Prepared Under Grant Nsg-490  
National Aeronautics and Space Administration

NASA CR70862  
**ENGINEERING RESEARCH LABORATORIES**

COLLEGE OF ENGINEERING  
UNIVERSITY OF ARIZONA  
TUCSON, ARIZONA



FACILITY FORM 602

19882 70885

\_\_\_\_\_  
(ACCESSION NUMBER)

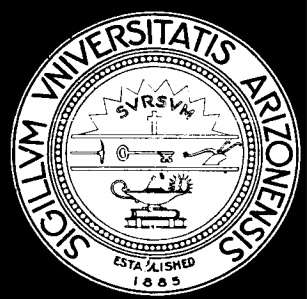
329  
(PAGES)

CR 70862  
(NASA CR OR TMX CR AD NUMBER)

\_\_\_\_\_  
(THRU)

1  
(CODE)

10  
(CATEGORY)



GPO PRICE \$ \_\_\_\_\_

CFSTI PRICE(S) \$ \_\_\_\_\_

Hard copy (HC) 7.00

Microfiche (MF) 1.75

SEMI-ANNUAL STATUS REPORT

RESEARCH IN AND APPLICATION OF MODERN AUTOMATIC CONTROL THEORY  
TO NUCLEAR ROCKET DYNAMICS AND CONTROL

Prepared Under Grant NsG-490  
National Aeronautics and Space Administration

BY

Lynn E. Weaver  
Donald G. Shultz  
James L. Melsa  
Philip A. Secker  
Donald M. Nordal  
Charles R. Slivinsky

Nuclear Engineering Department  
The University of Arizona  
Tucson, Arizona

February 1966

Engineering Research Laboratories  
The University of Arizona  
College of Engineering  
Tucson, Arizona

## PREFACE

This report is divided into three separate sections, each covering a different phase of the research grant.

In Section I, "Optimal Feedback Control of Nuclear Reactor Systems," Optimal Control Theory is applied in order to derive analytical expressions for compensating reactivity control which minimizes an integral quadratic performance index containing system errors and control motions. Various linear and nonlinear reactor models are analyzed. Analog computer studies show the effect of quasi-optimal feedback control in minimizing system errors caused by internal disturbances.

In the second part of this report, "Modeling with Liapunov Function," the Second Method of Liapunov is used to analyze the behavior of high-order control systems. This is accomplished by finding a lower order model whose response closely approximates the response of a higher order system. The model is developed by "matching" the surfaces described by Liapunov functions of the system and the model. In particular, a second-order model and a third-order model are developed which provide good results for all systems investigated. The second-order model is shown to be similar to the model obtained using phase margin techniques; and of greater importance, the third-order model is shown to be a better approximation to systems than the phase margin and the second order models. Thus this method serves to extend the practical usefulness of the Second Method from mere stability analysis to relative stability analysis (response of the system) and synthesis. Future effort will be directed to the problems of finding nonlinear models for nonlinear systems and specification of the accuracy of the model.

The third section, "Linear System Design Using State Variable Feedback," deals with the problems of the design of optimal feedback systems for linear system subject to quadratic integral performance criterion. In particular, two specific problems are attacked - the regulator problem and the servomechanism problem. In the regulator problem, the optimal design is shown to be a weighted constant feedback of all state variables. For the servomechanism problem, the solution consists of a regulator plus a linear prefitter system. Methods are presented for both the exact and approximate solution of both problems. Future work will be involved with the finite interval-of-control problem and the very important case when all of the state variables are not measurable.

✓

N66-19883

SECTION I

OPTIMAL FEEDBACK CONTROL OF NUCLEAR REACTOR SYSTEMS

## TABLE OF CONTENTS

	Page
LIST OF FIGURES . . . . .	v
LIST OF TABLES . . . . .	vii
ABSTRACT . . . . .	viii
INTRODUCTION AND OUTLINE . . . . .	1
Introduction . . . . .	1
Outline of the Thesis . . . . .	4
OPTIMAL CONTROL THEORIES . . . . .	7
Introduction and Notation . . . . .	7
Calculus of Variations . . . . .	12
Pontryagin's Maximum Principle . . . . .	24
Dynamic Programming . . . . .	27
DESIGN OF OPTIMAL FEEDBACK CONTROL SYSTEMS . . . . .	33
Introduction . . . . .	33
Feedback Control Design Considerations . . . . .	34
Synthesis of Linear Systems . . . . .	45
Synthesis of Non-Linear Quasi-Optimal Control Systems . . . . .	56
FIRST-ORDER NUCLEAR REACTOR SYNTHESIS . . . . .	64
Introduction . . . . .	64
Linear Prompt Neutron Reactor Synthesis . . . . .	65
Non-Linear Prompt Neutron Reactor Synthesis . . . . .	76
Linear Delayed Neutron Reactor Synthesis . . . . .	87
Non-Linear Delayed Neutron Reactor Synthesis . . . . .	91
Another Non-Linear Reactor Model . . . . .	95
HIGHER-ORDER NUCLEAR REACTOR SYNTHESIS . . . . .	100
Introduction . . . . .	100
Delayed Neutron Reactor Synthesis; Six and One Group Models . . . . .	102
Synthesis of Reactor Power Transfer . . . . .	110
Nuclear Rocket Engine Start-Up Synthesis . . . . .	122
REFERENCES . . . . .	138

## LIST OF FIGURES

		Page
Figure 1.	Electrical Network with Transfer Function Equivalent to Equation (3-9) . . . . .	40
Figure 2.	Schematic of Reactor with On-Line Computer	40
Figure 3.	Reactor System with a Pseudo-Control Variable . . . . .	46
Figure 4.	Optimal Feedback Controller . . . . .	53
Figure 5.	Optimal Feedback Gains for a Prompt-Reactor Model . . . . .	70
Figure 6.	Comparison of Power Responses for Different Feedback Criteria . . . . .	74
Figure 7.	Compensating Reactivity for Different Feedback Criteria . . . . .	74
Figure 8.	Optimal Control Reactivity with Minimum Period Constraint . . . . .	77
Figure 9.	Quasi-Optimal Control <u>vs</u> Exact Non-Linear Optimal . . . . .	85
Figure 10.	Feedback Controllers for Non-Linear Reactor Model . . . . .	86
Figure 11.	Ratio of Inverse Periods <u>vs</u> Allowable Power Deviation . . . . .	90
Figure 12.	Comparison of Control Laws for a Non-Linear Reactor with Delayed Neutrons . . . . .	94
Figure 13.	Exact Control for Different Power Coefficients of Reactivity . . . . .	97
Figure 14.	Comparison of Quasi-Optimal Controls with Exact for $\frac{\alpha n_0}{\beta} = 1.0$ . . . . .	99
Figure 15.	Sensitivity of Feedback Gain, $k_{11}$ , Due to $\ell$ . . . . .	109

LIST OF FIGURES--Continued

	Page
Figure 16. Open-Loop Nominal Reactivity . . . . .	112
Figure 17. Open-Loop Nominal Power . . . . .	113
Figure 18. Open-Loop Nominal Control . . . . .	114
Figure 19. Time-Varying Gains for Reactor Power Transfer . . . . .	118
Figure 20. Reactivity and Power for Reactor Power Transfer . . . . .	120
Figure 21. Nominal State Trajectories; $\gamma = 1.6$ . . .	124
Figure 22. Nominal Control Variables; $\gamma = 1.6$ . . .	125
Figure 23. Time-Varying Feedback Gains for Nuclear Rocket Engine . . . . .	131
Figure 24. Contaminated Open-Loop Controls; $\gamma = 1.6$ .	134
Figure 25. Effect of Feedback Control on Power . . . .	135
Figure 26. Effect of Feedback Control on Temperature .	136



LIST OF TABLES

	Page
TABLE I. Hypothetical Nuclear Rocket Engine Parameters	127

ABSTRACT

19883

Linear optimal feedback control theory is employed for the synthesis of several nuclear reactor models. Optimal feedback control theory is presented from the viewpoints of three commonly used techniques in modern control theory, namely: the calculus of variations, Pontryagin's maximum principle, and Bellman's dynamic programming. For the synthesis of linear nuclear reactor control problems, these three methods all yield identical optimal feedback controllers. For the synthesis of non-linear nuclear reactor models, approximation techniques based on either the maximum principle or dynamic programming are required, and the two methods yield different results.

The purpose of this study is to derive analytical expressions for a compensating reactivity control which minimizes an integral quadratic performance index containing system errors and control motions. First-order linear and non-linear reactor systems are analyzed in order to present as many facets of the optimal synthesis problem as possible using simple examples. Higher-than-first-order nuclear reactor models are analyzed for optimal linear and quasi-optimal non-linear cases. Analog computer studies show the effect of quasi-optimal feedback control in minimizing system errors caused by internal dynamic

disturbances. Time-varying feedback gain programs were determined by means of the digital computer for two different non-linear reactor models. The examples in this work demonstrate the usefulness of optimal feedback control synthesis for nuclear reactor systems.

*Author*

## Chapter 1

### INTRODUCTION AND OUTLINE

#### Introduction

In recent years a number of investigations have been carried out for the determination of optimal controls for nuclear reactor systems (3, 8, 11, 15). In the period since World War II, new mathematics of automatic control theory have been developed at a rapidly increasing rate. Linear control theory proved to be unsatisfactory for many types of dynamic control systems, which contain characteristic nonlinearities. Describing-function and phase-plane techniques were developed for non-linear systems. With the advent of analog and digital computers, very complicated control systems can be evaluated through indirect simulation. Most recently, time-domain control-system synthesis has begun to play a very important role in automatic control. To the engineer, the time-domain formulation is a realistic framework in which to work and affords a wider range of problems which may be handled.

This thesis is concerned primarily with the problems of optimization of feedback control systems. The mathematics of optimization in automatic control are probably the most notable contributions to control theory of any. The idea of optimization is surely not new and stems directly from

the classical mathematics of the calculus of variations. However, the application of mathematical optimization in automatic control is a significant divergence from the methods discussed in the previous paragraph. Optimization theory attempts to yield a system controller which exactly incorporates all of the control system design constraints directly without the need for trial-and-error system analysis. The recent theories of optimization used most frequently in control system design are those of Pontryagin (10) and Bellman (1). Both of these theories give necessary and sufficient conditions for the optimal control of linear dynamic systems and as expected the resulting controls are identical. For non-linear systems Pontryagin's maximum principle gives necessary but not sufficient conditions for the optimal control. On the contrary, however, Bellman's dynamic programming approach gives both necessary and sufficient conditions for non-linear optimal control systems also. For other than linear systems with unbounded control regions, these theories involve sophisticated and complex computational techniques.

The application of optimization techniques to reactor dynamics is new and almost all studies in this area have been published since 1961. One of the very first studies of optimal processes in nuclear engineering was done by Rosztoczy (11,12). Since then, other studies have been continuously forthcoming. The general trend in all but

a few of these studies (5, 6) has been to obtain an open-loop optimal control, for specific reactor systems, using fixed performance criteria. In other words, the resulting control law is satisfactory, only (a) if the reactor model exactly (mathematically) represents the fixed physical reactor, (b) for one set of initial conditions, and (c) where no disturbances occur. Due to these practical restrictions the usefulness of such a control is questionable.

Thus a practical need is established for a closed-loop feedback control. In this work, the optimization methods of Pontryagin's maximum principle and Bellman's dynamic programming are applied to a wide range of reactor kinetic problems to determine optimal feedback controllers. The primary effort is to establish a compensating reactivity feedback controller which minimizes, in an optimal sense, deviations of the instantaneous reactor states (for example neutron density and delayed neutron precursor densities) from the desired or nominal states. Linear and non-linear reactor systems are considered. In those cases where the reactor dynamics considered are non-linear, approximate methods of determining the optimal feedback control are used. The resulting controller is nearly optimal, hereafter termed "quasi-optimal", and increases in component complexity as greater accuracy is required.

The usefulness of optimal feedback control for nuclear reactors of all types is demonstrated in this work. Current optimal control theories are developed to

the extent that both deterministic and statistical control systems can be handled (4). Statistical control systems are characterized by uncertainties in measurements and/or sporadic fluctuations of the physical states of the system. At present only a limited class of problems can be evaluated using statistical, or stochastic, optimal control theory. Only deterministic, or exactly measurable, state variable systems are considered here.

### Outline of the Thesis

There are essentially two distinct parts in this work. Chapters 1, 2, and 3 are of an introductory nature and chapters 4, and 5 contain examples of optimal feedback control theory applied to specific reactor systems.

Chapter 2 discusses the mathematical theories of optimal processes necessary to formulate and ultimately to solve the feedback controller equations. Brief discussions are contained in this chapter on the calculus of variations, Pontryagin's maximum principle, and Bellman's dynamic programming. Only generalizations, including necessary and sufficient conditions for optimization, are discussed.

In chapter 3 the specialization of optimal control theory to the feedback control problem is considered. Discussions of such topics as the selection of appropriate performance criteria, control system stability, weighting factor selection, and control system constraints are included here. In addition the distinction between finite

and infinite control intervals are discussed. Most of the background material necessary for actual evaluation of the control system is contained in this chapter. Of greatest importance are the approximation techniques required when applying the maximum principle or dynamic programming to non-linear systems: i.e. in non-quadratic error criterion, feedback-control synthesis.

A number of linear and non-linear first-order reactor dynamic examples are presented in chapter 4. The emphasis here is to demonstrate and compare, in a simple way, the techniques developed. An effort is made to reveal the complexity of the synthesis problem even for first-order systems, thus giving some insight into the complexities of higher-order, non-linear systems. All reactor examples in this chapter are based on a steady state desired power level (neutron density) and an infinite control interval. It is felt that optimal feedback control for ordinary power reactors, operating in the steady state, is demonstrated in this chapter.

In chapter 5 optimal feedback-control theory is extended to reactor-dynamic systems of order greater than first operating in other than the steady state condition. First, a linear example of a reactor with delayed neutrons is presented. The purpose here is to obtain constant fixed optimal feedback gains for a reactor with six groups of delayed neutrons that can be used for any reactor



straightforwardly. A comparison is made with a one delayed neutron group model. Next, Pontryagin's maximum principle is applied to obtain a quasi-optimal control for a non-linear reactor model which undergoes a power increase from 10 to 50 kilowatts in a finite time. Finally, a quasi-optimal feedback control is determined for the startup of a nuclear rocket engine. In this example two control variables are optimized, discontinuities in nominal controls are considered, and a finite control interval is used. Analog computer simulation studies show the effectiveness of quasi-optimal control clearly.

## Chapter 2

### OPTIMAL CONTROL THEORIES

#### Introduction and Notation

In this chapter the various theories of mathematical optimization are presented as the basis for optimal feedback control system design. Three separate methods of functional optimization are described; the calculus of variations (2), Pontryagin's maximum principle (10), and Bellman's dynamic programming(1).

The similarities of the calculus of variations and the maximum principle are many. The calculus of variations, however, is more restrictive than the maximum principle in the types of variational problems that can be handled. The maximum principle extends the classical mathematics of the calculus of variations to include solutions of problems with algebraic inequalities. Both theories have been proven to be necessary and sufficient conditions for optimization of linear problems and require the indirect solution of multi-point boundary-value problems.

For non-linear problems both are a necessary, but not sufficient, condition for optimization. The basic theorems of the calculus of variations are first discussed. The maximum principle is briefly outlined later, with the main differences clearly emphasized.

In 1957, Bellman (1) formulated the theory of dynamic programming. Whereas the maximum principle and the calculus of variations are classified as indirect theories, because the solution of two-point boundary-value problems are required, dynamic programming is classified as a direct method of optimization. In the discrete form of this approach a single problem in  $N$  variables is transformed into  $N$  problems, each in one variable, and a direct search for the optimal "policy," or solution, is required. Conceptually, this is a far easier task than the solution of a multi-point boundary-value problem. However, the number of computations required for a final solution roughly increases exponentially with the order of the problem, greatly restricting the solution of any sizable problem. One very important aspect of dynamic programming is that equality or inequality constraints on the problem reduce the regions of search for the optimal policy and in principle simplify the solution. This is in direct contrast with the maximum principle.

The generalized optimization problem can be expressed fairly simply. Since dynamic control systems are the only type of problem considered here, the notation and presentation is kept in accord with recent control and nuclear reactor literature.

In control system design, the mathematics of the device to be controlled are usually given. This

mathematical description is called the dynamic process. In this work the dynamic processes are nuclear reactors. The inputs of the dynamic process are termed control variables and are designated by the quantities  $u_1(t)$ ,  $u_2(t)$ , ...,  $u_M(t)$ , but in specific reactor examples the quantity  $\rho(t)$ , control reactivity, is sometimes used. The dependent variables, or outputs, are termed the response variables  $q_1(t), q_2(t), \dots, q_Q(t)$ . These response variables may not always represent the physical variables of the dynamic process, but in all cases are functionally related to these variables. The actual physical outputs of the dynamic process are termed state variables  $x_1(t), x_2(t), \dots, x_N(t)$ . In the dynamic process of a reactor described by six or more groups of delayed neutrons, where only the neutron density is measurable, the response variable would be the neutron density and the state variables would be the neutron density together with all the delayed neutron precursor densities. The minimum number of state variables which completely describe the dynamic process, a set of first-order ordinary differential equations, is equal to the order  $N$  of the system.

The optimal control problem is then defined as the problem of controlling the dynamic process in such a way that the performance of the system is optimum according to some specified functional performance criterion. Not only must the control optimize the performance criterion, but it must do so without violating any of the physical

constraints to which the dynamic process is subjected.

Using column vector notation, the control, response, and state variables are represented as

$$\underline{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \cdot \\ \cdot \\ u_M(t) \end{bmatrix} ; \quad \underline{q}(t) = \begin{bmatrix} q_1(t) \\ q_2(t) \\ \cdot \\ \cdot \\ q_Q(t) \end{bmatrix} ; \quad \underline{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdot \\ \cdot \\ x_N(t) \end{bmatrix} \quad (2-1)$$

In general the differential equations describing the nuclear reactor dynamic processes are non-linear with time-varying coefficients and can be represented as

$$\frac{d[\underline{x}(t)]}{dt} = \dot{\underline{x}}(t) = \underline{f}[\underline{x}(t), \underline{u}(t), t] \quad (2-2)$$

which is identical with the set of first-order differential equations

$$\dot{x}_1(t) = f_1[\underline{x}(t), \underline{u}(t), t] \quad i = 1, 2, \dots, N \quad (2-3)$$

The physical design constraints, or saturation constraints, on the control and state variables of the dynamic process are

$$\underline{u}(t) \in U(t) \text{ and } \underline{x}(t) \in X(t) \quad (2-4)$$

where the notation  $\underline{u}(t) \in U(t)$  designates that the vector  $\underline{u}(t)$  lies within, or on the boundary of, the closed region  $U(t)$  of the control vector space.

When the dynamic process is linear equation (2-2) can be specialized by the notation

$$\dot{\underline{x}}(t) = A(t)\underline{x}(t) + B(t)\underline{u}(t) \quad (2-5)$$

If the dynamic process is linear, saturation does not occur and equations (2-4) are eliminated. The time-varying matrices  $A(t)$  and  $B(t)$  are written in the form

$$A(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) & \dots & a_{1N}(t) \\ a_{21}(t) & a_{22}(t) & \dots & a_{2N}(t) \\ \vdots & \vdots & & \vdots \\ a_{N1}(t) & a_{N2}(t) & \dots & a_{NN}(t) \end{bmatrix} \quad (2-6)$$

$$B(t) = \begin{bmatrix} b_{11}(t) & b_{12}(t) & \dots & b_{1M}(t) \\ b_{21}(t) & b_{22}(t) & \dots & b_{2M}(t) \\ \vdots & \vdots & & \vdots \\ b_{N1}(t) & b_{N2}(t) & \dots & b_{NM}(t) \end{bmatrix}$$

The performance criterion, that must be satisfied for optimal control, is of prime importance and must be selected carefully and realistically. The instantaneous performance criterion,  $e_m$ , is calculated in terms of the function

$$e_m(t) = h[\underline{q}(t), \underline{u}(t), t] \quad (2-7)$$

The total system performance over the present and future

time,  $t \leq \tau \leq T$ , during which control effort is applied to the system is found by integrating equation (2-7)

$$e(t) = \int_t^T h[\underline{q}(\tau), \underline{u}(\tau), \tau] d\tau \quad (2-8)$$

The response variables are always functions of the physical state variables and equations (2-7) and (2-8) are rewritten

$$e_m(t) = H[\underline{x}(t), \underline{u}(t), t] \quad (2-9)$$

$$\text{and } e(t) = \int_t^T H[\underline{x}(\tau), \underline{u}(\tau), \tau] d\tau \quad (2-10)$$

### Calculus of Variations

Three problems were responsible for the development of the calculus of variations (2). The brachistochrone problem is the simplest and involves determining a curve between two fixed end-points such that a particle sliding along the curve under the influence of gravity travels between the end-points in minimum time. From this problem evolved the basic conditions for the minimization of a functional equation with no constraints.

The problem of geodesics is concerned with minimizing a functional equation subject to a finite constraint. For example, it may be desired to find the curve of minimal length lying on a given surface joining two fixed points on that surface.

Finally, the isoperimetric problem is concerned with finding a closed curve, of given length, such that

the enclosed surface area is maximum. This is the minimization of a functional, subject to an integral constraint.

The basic equation that gives the necessary condition for a maximum, or minimum, of an integral functional is the Euler-Lagrange equation. Equation (2-10), here repeated, is an example of the integral functional considered in this study.

$$e(t) = \int_t^T H[\underline{x}(\tau), \underline{u}(\tau), \tau] d\tau \quad (2-11)$$

In deriving the Euler-Lagrange necessary condition, equation (2-11) will be used in a modified form. Equation (2-2) relates the control vector,  $\underline{u}(t)$ , to the state variable vector,  $\underline{x}(t)$ , and its derivative,  $\dot{\underline{x}}(t)$ . In other words

$$\underline{u}(t) = \underline{g}[\underline{x}(t), \dot{\underline{x}}(t), t] \quad (2-12)$$

and upon substitution equation (2-11) becomes

$$e(t) = \int_t^T F[\underline{x}(\tau), \dot{\underline{x}}(\tau), \tau] d\tau \quad (2-13)$$

The problem of minimizing this performance index is the problem most frequently treated and is equivalent to the brachistochrone problem when  $\underline{x}(t)$  is a first-order vector, i.e. the scalar  $x(t)$ . In order to simplify the derivation of the Euler-Lagrange equation, a first-order process is considered initially.

$$e(t) = \int_t^T F[x(\tau), \dot{x}(\tau), \tau] d\tau \quad (2-14)$$



The minimization of equation (2-24) is performed by first assuming that the state variable is

$$x(\tau) = x^*(\tau) + \alpha \delta x(\tau) \quad (2-15)$$

where  $x^*(\tau)$  is the function that actually minimizes equation (2-14). Here  $\alpha$  is an arbitrarily small quantity and  $\delta x(\tau)$  is considered to be an arbitrary and unrestricted perturbation. The derivative of  $x(\tau)$  is written

$$\dot{x}(\tau) = \dot{x}^*(\tau) + \alpha \delta \dot{x}(\tau) \quad (2-16)$$

Since  $\alpha$  is arbitrarily small the performance index is perturbed infinitesimally about  $F[x^*(\tau), \dot{x}^*(\tau), \tau]$ . If  $F$  and its derivatives with respect to  $x(\tau)$  and  $\dot{x}(\tau)$  are continuous, the instantaneous performance criterion is expanded in a Taylor series such that

$$\begin{aligned} F[x(\tau), \dot{x}(\tau), \tau] &= F[x^*(\tau), \dot{x}^*(\tau), \tau] + \alpha \left\{ \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial x^*(\tau)} \delta x(\tau) \right. \\ &\quad \left. + \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)} \delta \dot{x}(\tau) \right\} + \alpha^2 \{ \dots \} + \dots \end{aligned} \quad (2-17)$$

where  $\alpha^2$  is multiplied by all second partial derivatives of  $F^*$  with respect to  $x^*$  and  $\dot{x}^*$  and where the notation  $\frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial x^*(\tau)}$  is equivalent to

$$\frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial x^*(\tau)} \equiv \frac{\partial F[x(\tau), \dot{x}(\tau), \tau]}{\partial x(\tau)} \bigg|_{\substack{x=x^* \\ \dot{x}=\dot{x}^*}} \quad (2-18)$$

If the above series converges uniformly, the performance index can be written

$$e(t) = e^*(t) + \int_t^T \alpha \left\{ \delta x(\tau) \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial x^*(\tau)} + \delta \dot{x}(\tau) \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)} \right\} d\tau + \alpha^2 \{ \dots \} + \dots \quad (2-19)$$

where  $e^*(t)$  is the performance index evaluated at  $x(\tau) = x^*(\tau)$  which minimizes  $e(t)$ .

The first necessary condition for a minimum is

$$\left. \frac{\partial e(t)}{\partial \alpha} \right|_{\alpha=0} = 0 \quad (2-20)$$

when  $\delta x(\tau)$  is an arbitrary function. The result of this condition applied to equation (2-19) is

$$\int_t^T \left\{ \delta x(\tau) \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial x^*(\tau)} + \delta \dot{x}(\tau) \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)} \right\} d\tau = 0 \quad (2-21)$$

Integrating the second term in this expression by parts

$$\begin{aligned} \int_t^T \delta \dot{x}(\tau) \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)} d\tau &= \\ - \int_t^T \delta x(\tau) \frac{d}{d\tau} \left\{ \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)} \right\} d\tau & \quad (2-22) \\ + \delta x(\tau) \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)} \Big|_{\tau=t}^{\tau=T} & \end{aligned}$$

Equation (2-21) now becomes

$$\int_t^T \delta x(\tau) \left[ \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial x^*(\tau)} - \frac{d}{d\tau} \left\{ \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)} \right\} \right] d\tau + \delta x(\tau) \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)} \Big|_{\tau=t}^{\tau=T} = 0 \quad (2-23)$$

If the integrand of equation (2-23) is finite at  $\tau = t$  and  $\tau = T$  then the contribution at these end-points is due only to the 2<sup>nd</sup> term. This is the so-called transversality condition

$$\delta x(\tau) \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)} \Big|_{\tau=t}^{\tau=T} = 0 \quad (2-24)$$

Since  $\delta x(\tau)$  is arbitrary, the integrand of equation (2-23) must vanish independently of  $\delta x(\tau)$  on the interval  $t < \tau < T$ . This is the Euler-Lagrange necessary condition for optimality.

$$\frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial x^*(\tau)} - \frac{d}{d\tau} \left\{ \frac{\partial F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)} \right\} = 0 \quad (2-25)$$

The appropriate boundary conditions are required for explicit solutions of equation (2-25). If these boundary conditions are specified as  $x^*(t)$  and  $\dot{x}^*(t)$  or as  $x^*(T)$  and  $\dot{x}^*(T)$  the solution is the common initial-value or one-point boundary-value problem. If, however, they are specified as  $x^*(t)$  and  $x^*(T)$  or as  $x^*(t)$  and  $\dot{x}^*(T)$  then the solution is termed a two-point boundary-value problem.

The initial state of the dynamic process is always fixed as

$$x(t) = x^*(t) \quad (2-26)$$

which requires that  $\delta x(t) = 0$  for the transversality condition of equation (2-24). The minimizing function  $x^*(\tau)$  must have a fixed-point boundary condition, and because  $\delta x(t) = 0$ , from equation (2-24)

$$\delta x(T) \frac{\partial F[x^*(T), \dot{x}^*(T), T]}{\partial \dot{x}^*(T)} = 0 \quad (2-27)$$

If this fixed-point boundary condition is

$$x(T) = x^*(T) \quad (2-28)$$

then  $\delta x(T) = 0$  and equation (2-27) is satisfied automatically. If, however,  $x^*(T)$  is free to assume any finite value then the so-called free-point terminal-boundary condition results and

$$\frac{\partial F[x^*(T), \dot{x}^*(T), T]}{\partial \dot{x}^*(T)} = 0 \quad (2-29)$$

because  $\delta x(T)$  is arbitrary.

In order that the instantaneous performance index be a minimum value, one additional condition must be satisfied. This is known as the Legendre condition and is

$$\left. \frac{\partial^2 e(t)}{\partial \alpha^2} \right|_{\alpha=0} > 0 \quad (2-30)$$

This is seen to be equivalent to the minimum of a function given by differential calculus. The application of this condition to equation (2-19), extended to include the product of  $\alpha^2$  times the 2<sup>nd</sup> derivative terms, gives

$$\begin{aligned}
& \frac{1}{2} \int_t^T \left\{ \delta x(\tau)^2 \frac{\partial^2 F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial x^*(\tau)^2} \right. \\
& + 2 \delta x(\tau) \delta \dot{x}(\tau) \frac{\partial^2 F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial x^*(\tau) \partial \dot{x}^*(\tau)} \\
& \left. + \delta \dot{x}(\tau)^2 \frac{\partial^2 F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)^2} \right\} d\tau > 0
\end{aligned} \tag{2-31}$$

A sufficient condition for satisfying equation (2-31) everywhere on  $t \leq \tau \leq T$  is a positive integrand for any  $\delta x(\tau)$  and  $\delta \dot{x}(\tau)$ . A positive integrand is ensured when

$$\left| \begin{array}{cc}
\frac{\partial^2 F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial x^*(\tau)^2} & \frac{\partial^2 F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial x^*(\tau) \partial \dot{x}^*(\tau)} \\
\frac{\partial^2 F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau) \partial x^*(\tau)} & \frac{\partial^2 F[x^*(\tau), \dot{x}^*(\tau), \tau]}{\partial \dot{x}^*(\tau)^2}
\end{array} \right| > 0 \tag{2-32}$$

The sufficient condition of equation (2-32) is very restrictive and it is difficult, if not impossible, to test a given performance criterion for sufficiency. For these reasons some authors (7) tacitly assume that the Euler-Lagrange necessary condition of equation (2-25) is both a necessary and a sufficient condition for minimization. Functions which satisfy equation (2-32) are given the term, strictly convex functions.

There are many solutions of the Euler-Lagrange equation which are integrable, but one such problem is of special interest in the solution of the exact non-linear optimal control for first-order dynamic processes. This

solution will be considered here because of its frequent application to the non-linear problems of chapter 4.

If the instantaneous performance criterion,  $F$ , depends on  $x(t)$  and  $\dot{x}(t)$  only and is independent of  $t$ , the following solution results.

$$F = F[x(t), \dot{x}(t)] \quad (2-33)$$

The Euler-Lagrange equation can be shown to be

$$\begin{aligned} \frac{\partial F[x^*(\tau), \dot{x}^*(\tau)]}{\partial x^*(\tau)} - \frac{\partial^2 F[x^*(\tau), \dot{x}^*(\tau)]}{\partial x^*(\tau) \partial \dot{x}^*(\tau)} \dot{x}^*(\tau) \\ - \frac{\partial^2 F[x^*(\tau), \dot{x}^*(\tau)]}{\partial \dot{x}^*(\tau)^2} \ddot{x}^*(\tau) = 0 \end{aligned} \quad (2-34)$$

When both sides of equation (2-34) are multiplied by the function  $\dot{x}^*(t)$  the Euler-Lagrange equation becomes the exact derivative

$$0 = \frac{d}{d\tau} \left\{ F[x^*(\tau), \dot{x}^*(\tau)] - \dot{x}^*(\tau) \frac{\partial F[x^*(\tau), \dot{x}^*(\tau)]}{\partial \dot{x}^*(\tau)} \right\} \quad (2-35)$$

Consequently, the Euler-Lagrange equation has the first integral

$$\left\{ F[x^*(\tau), \dot{x}^*(\tau)] - \dot{x}^*(\tau) \frac{\partial F[x^*(\tau), \dot{x}^*(\tau)]}{\partial \dot{x}^*(\tau)} \right\} = C \quad (2-36)$$

where  $C$  is the constant of integration. Thus the exact optimal solution for  $\dot{x}^*(\tau)$  is

$$\dot{x}^*(\tau) = \frac{F[x^*(\tau), \dot{x}^*(\tau)] - C}{\frac{\partial F[x^*(\tau), \dot{x}^*(\tau)]}{\partial \dot{x}^*(\tau)}} \quad (2-37)$$

Equation (2-12) relates the optimal control of this first-order dynamic process to the above equation.

The generalized extension of the calculus of variations to  $N^{\text{th}}$ -order dynamic processes is presented in terms of Lagrange multipliers. The results directly follow the previous development. The problem is left in terms of minimizing the original functional equation (2-10), here repeated,

$$e(t) = \int_t^T H[\underline{x}(\tau), \underline{u}(\tau), \tau] d\tau \quad (2-38)$$

which is subject to the constraints of the dynamic process

$$\dot{\underline{x}}(t) = \underline{f}[\underline{x}(t), \underline{u}(t), t] \quad (2-39)$$

This is the so-called isoperimetric problem with integral constraints and is reformulated in terms of a constrained performance criterion, which incorporates both equations (2-38) and (2-39). It is written here in terms of the Lagrange multipliers

$$e_c(t) = \int_t^T H_c[\underline{x}(\tau), \underline{u}(\tau), \underline{\lambda}(\tau), \tau] d\tau \quad (2-40)$$

where  $\underline{\lambda}(\tau)$  is the Lagrange multiplier vector. There is a Lagrange multiplier for each state variable. The constrained instantaneous performance measure is

$$H_c[\underline{x}(\tau), \underline{u}(\tau), \underline{\lambda}(\tau), \tau] = H[\underline{x}(\tau), \underline{u}(\tau), \tau] \quad (2-41)$$

$$+ \sum_{n=1}^N \lambda_n(\tau) \left\{ f_n[\underline{x}(\tau), \underline{u}(\tau), \tau] - \dot{x}_n(\tau) \right\}$$

From this equation it is seen that the value of the constrained performance criterion is equal to the original performance criterion when the Lagrange multipliers,  $\lambda_n(\tau)$  ( $n = 1, 2, \dots, N$ ), are chosen such that the terms inside the braces vanish.

The conditions for a minimum are developed as before with the inclusion of the perturbed optimal control variables and perturbed optimal Lagrange multipliers.

$$x_1(\tau) = x_1^*(\tau) + \alpha_1 \delta x_1(\tau) \quad (2-42)$$

$$\dot{x}_1(\tau) = \dot{x}_1^*(\tau) + \alpha_1 \delta \dot{x}_1(\tau) \quad (2-43)$$

$$\lambda_1(\tau) = \lambda_1^*(\tau) + \beta_1 \delta \lambda_1(\tau) \quad (2-44)$$

$$u_j(\tau) = u_j^*(\tau) + \gamma_j \delta u_j(\tau) \quad j = 1, 2, \dots, M \quad (2-45)$$

where  $M$  represents the number of control inputs to the dynamic process. Now the constrained performance criterion  $H_c$  is expanded in a Taylor series about the optimal value  $H_c^*$ . At this point all arguments of the functions are dropped for conciseness.

$$H_c = H_c^* + \sum_{i=1}^N \left\{ \alpha_i \delta x_i \frac{\partial H_c^*}{\partial x_i^*} + \alpha_i \delta \dot{x}_i \frac{\partial H_c^*}{\partial \dot{x}_i^*} + \beta_i \delta \lambda_i \frac{\partial H_c^*}{\partial \lambda_i^*} \right\} \quad (2-46)$$

$$+ \sum_{j=1}^M \gamma_j \delta u_j \frac{\partial H_c^*}{\partial u_j^*} + \dots$$



All perturbations are treated independently and the conditions for a minimum are

$$\frac{\partial e_c(t)}{\partial \alpha_1} = 0 \quad ; \quad \frac{\partial e_c(t)}{\partial \beta_1} = 0 \quad ; \quad \frac{\partial e_c(t)}{\partial \gamma_j} = 0 \quad (2-47)$$

where these derivatives are all evaluated at  $\underline{\alpha} = \underline{\beta} = \underline{\gamma} = \underline{0}$ .

When these conditions are applied to equations (2-40) and (2-46) the result is

$$\int_t^T \left\{ \delta x_1 \left[ \frac{\partial H_c^*}{\partial x_1^*} \right] + \delta \dot{x}_1 \left[ \frac{\partial H_c^*}{\partial \dot{x}_1^*} \right] \right\} d\tau = 0 \quad i = 1, 2, \dots, N \quad (2-48)$$

$$\int_t^T \delta \lambda_1 \left[ \frac{\partial H_c^*}{\partial \lambda_1^*} \right] d\tau = 0 \quad i = 1, 2, \dots, N \quad (2-49)$$

$$\int_t^T \delta u_j \left[ \frac{\partial H_c^*}{\partial u_j^*} \right] d\tau = 0 \quad j = 1, 2, \dots, M \quad (2-50)$$

Integrating the second term of equation (2-48) by parts and combining the results with the first term the result is

$$\int_t^T \delta x_1 \left\{ \frac{\partial H_c^*}{\partial x_1^*} - \frac{d}{d\tau} \left[ \frac{\partial H_c^*}{\partial \dot{x}_1^*} \right] \right\} d\tau + \delta x_1 \left[ \frac{\partial H_c^*}{\partial \dot{x}_1^*} \right] \Bigg|_{\tau=t}^{\tau=T} = 0 \quad i = 1, 2, \dots, N \quad (2-51)$$

The result, as in the first-order case, is a set of Euler-Lagrange equations and transversality conditions corresponding to each of the N state variables. They are summarized as

$$\delta x_1 \frac{\partial H_c^*}{\partial \dot{x}_1^*} \bigg|_{\tau=t}^{\tau=T} = 0 \quad i = 1, 2, \dots, N \quad (2-52)$$

$$\text{and} \quad \frac{\partial H_c^*}{\partial x_1^*} - \frac{d}{d\tau} \left[ \frac{\partial H_c^*}{\partial \dot{x}_1^*} \right] = 0 \quad i = 1, 2, \dots, N \quad (2-53)$$

In addition an Euler-Lagrange equation results for each of the N Lagrange multipliers and M control variables for arbitrary perturbations. Equations (2-49) and (2-50) give these necessary conditions as

$$\frac{\partial H_c^*}{\partial \lambda_1^*} = \dot{x}_1^* \quad i = 1, 2, \dots, N \quad (2-54)$$

$$\text{and} \quad \frac{\partial H_c^*}{\partial u_j^*} = 0 \quad j = 1, 2, \dots, M \quad (2-55)$$

For the fixed-point boundary conditions of

$$x_1^*(t) = x_1(T) \quad (2-56)$$

the perturbations  $\delta x_1(\tau)$  are zero and equation (2-52) is automatically satisfied. However, for the free-point boundary conditions where  $x_1^*(T)$  are allowed to assume any finite value at this end-point the boundary conditions are

$$x_1^*(t) = x_1(t) \quad ; \quad \lambda_1^*(T) = 0 \quad (2-57)$$

From equation (2-55) the optimal control is seen to be an algebraic relation and when substituted into the differential

equations of the dynamic process and the Lagrange multipliers give a set of  $2N$  first-order differential equations subject to the boundary conditions previously specified.

The calculus of variations can treat problems with movable boundaries, extremals with cusps (discontinuities within  $t < \tau < T$ ), and others, but the conditions outlined here are basic to any of the variational problems of this method.

### Pontryagin's Maximum Principle

The primary limitation of the calculus of variations in control theory is that the theory, as developed, is not suitable for solving problems where control, or state, variable saturation occurs. Pontryagin and his co-workers have extended the methods of calculus of variations to include such problems. Briefly the equations of Pontryagin will be outlined here using the Hamiltonian formulation.

The Hamiltonian function  $H'$  is related to the constrained instantaneous performance criterion of equation (2-39) and is stated here

$$H'[\underline{x}(\tau), \underline{u}(\tau), \underline{\lambda}(\tau), \tau] = \sum_{n=0}^N \lambda_n(\tau) f_n[\underline{x}(\tau), \underline{u}(\tau), \tau] \quad (2-58)$$

where  $\lambda_0(\tau) = 1 \quad (2-59)$

and  $\dot{x}_0(\tau) = f_0[\underline{x}(\tau), \underline{u}(\tau), \tau] = H[\underline{x}(\tau), \underline{u}(\tau), \tau] \quad (2-60)$

Thus  $H'$  is very nearly  $H_c$  of equation (2-39) but does not

include the derivatives of the state variables. The conditions for a minimum are

$$\frac{\partial H'}{\partial x_1^*} = -\dot{\lambda}_1^* \quad i = 1, 2, \dots, N \quad (2-61)$$

$$\frac{\partial H'}{\partial \lambda_1^*} = \dot{x}_1^* \quad i = 1, 2, \dots, N \quad (2-62)$$

$$\text{and } \frac{\partial H'}{\partial u_j^*} = 0 \quad j = 1, 2, \dots, M \quad (2-63)$$

These equations are seen to be equivalent to the Euler-Lagrange equations of the calculus of variations. Equation (2-63) is the result where saturation of the control variables does not occur. If the control vector must remain on or within a closed region of the M-dimensional control space U, the minimization process with respect to the control vector is written

$$H'^* = \min_{\underline{u}(\tau) \in U} H'[\underline{x}(\tau), \underline{u}(\tau), \underline{\lambda}(\tau), \tau] \quad (2-64)$$

Using the definition of H' the result of this minimization gives the following conditions for optimality

$$\frac{\partial H'^*}{\partial u_j^*} \frac{\partial u_j^*}{\partial x_1} = \frac{\partial H'^*}{\partial u_j^*} \frac{\partial u_j^*}{\partial \lambda_1} = 0 \quad (2-65)$$

since  $\partial H'^*/\partial u_j^* = 0$  when  $u_j^*$  is not on the boundary of U, and  $\partial u_j^*/\partial x_1 = \partial u_j^*/\partial \lambda_1 = 0$  when  $u_j^*$  is on the boundary of U.

The variables  $\lambda_1(t)$  in Pontryagin's maximum principle are sometimes termed Pontryagin variables, or adjoint variables, but are seen to be identical with the Lagrange multipliers of the calculus of variations.

In the special case where the dynamic process and the instantaneous performance criterion are time-invariant the Hamiltonian is independent of time and is written

$$H'^* = H'[\underline{x}(\tau), \underline{u}^*(\tau), \underline{\lambda}(\tau)] \text{ such that} \quad (2-66)$$

$$\frac{\partial H'^*}{\partial \tau} = 0$$

and the Hamiltonian is seen to be the constant of integration

$$H'^* = C \quad (2-67)$$

Also, in this case, the following condition for optimal control results

$$\dot{u}_j^* \frac{\partial H'^*}{\partial u_j^*} = 0 \quad j = 1, 2, \dots, M \quad (2-68)$$

since  $\partial H'^*/\partial u_j^* = 0$  for  $u_j^*$  not on the boundary of  $U$  and  $\dot{u}_j^* = 0$  for  $u_j^*$  on the boundary, because of the time invariance of the process.

The determination of the optimal control for a first-order dynamic process with control variable saturation is a relatively simple exercise. This is demonstrated in the several first-order examples of chapter 4. In addition, when no weight is placed on control effort in the performance

criterion, i.e. the performance criterion is not an explicit function of the vector  $\underline{u}(\tau)$ , bang-bang control generally results as the optimal for the case with saturation.. Unfortunately for higher-than-first-order dynamic processes where saturation occurs, the determination of the optimum control is a far more difficult problem to cope with. This is because differential equations, rather than algebraic relations, must be satisfied at all of the switching times.

### Dynamic Programming

Dynamic programming, as stated, is important in control theory for two reasons. Problems with control variable saturation are solvable, and the solution of the two-point boundary-value problem is not required. Control saturation can be handled since the assumption of unrestricted control perturbations is not made. The two-point boundary-value problem is eliminated by a flooding procedure where the optimal control signal is constructed point-by-point. Merriam (7) has stated this flooding procedure as follows:

" ... Dynamic programming embeds the solution to the optimization of the control system for a particular state of the dynamic process into the optimization of the control system for all possible states of the dynamic process. ..."

Thus one N-dimensional problem becomes N one-dimensional problems for all initial states and the appropriate optimal control is chosen from the resulting solutions.

In the derivation of the dynamic programming equations, again a first-order dynamic process will be used for simplicity. The functional desired to be optimized is that of equation (2-14), here repeated

$$e(t) = \int_t^T F[x(\tau), \dot{x}(\tau), \tau] d\tau \quad (2-69)$$

The initial concept in dynamic programming is that, rather than determining the optimal state variable  $x^*(\tau)$ , the minimization is determined by finding the optimal  $\dot{x}^*(\tau)$ . Thus a class of solutions is obtained and the particular  $x^*(t')$  is determined from the initial state and the value  $t'$  by the relation

$$x^*(t') = x(t) + \int_t^{t'} \dot{x}^*(\tau) d\tau \quad (2-70)$$

From this last equation it is seen that  $F$  then is simply a function of  $x(t)$ ,  $t$  and  $\tau$ , so that equation (2-69) is written

$$e^*(t) = \int_t^T F_m[x(t), t, \tau] d\tau \quad (2-71)$$

and thus the minimum value of the performance criterion is only dependent on  $x(t)$  and  $t$ .

$$e^*(t) = E[x(t), t] \quad (2-72)$$

The function  $E$  then can be called the minimum performance criterion, and is restated as

$$E[x(t), t] = \min_{\dot{x}(\tau)} \int_t^T F[x(\tau), \dot{x}(\tau), \tau] d\tau \quad (2-73)$$

When the derivative is restricted to remain in a given region, which is the case of control saturation, equation (2-73) is written

$$E[x(t), t] = \min_{\dot{x}(\tau) \in S(\tau)} \int_t^T F[x(\tau), \dot{x}(\tau), \tau] d\tau \quad (2-74)$$

From this equation it is seen that the boundary condition on the minimum performance criterion is that

$$E[x(T), T] = 0 \quad (2-75)$$

Using equation (2-74) the minimization problem can be restated as

$$\min_{\dot{x}(\tau) \in S(\tau)} \left\{ \int_{t'}^T F[x(\tau), \dot{x}(\tau), \tau] d\tau - E[x(t'), t'] \right\} = 0 \quad (2-76)$$

The integral within the braces can be broken into two separate integrals;

$$\min_{\dot{x}(\tau) \in S(\tau)} \left\{ \int_{t'}^{t'+\delta} F[x(\tau), \dot{x}(\tau), \tau] d\tau + \int_{t'+\delta}^T F[x(\tau), \dot{x}(\tau), \tau] d\tau - E[x(t'), t'] \right\} = 0 \quad (2-77)$$

Now using the definition of the minimum performance criterion equation (2-77) becomes



$$\min_{\dot{x}(\tau) \in S(\tau)} \left\{ \int_{t'}^{t'+\delta} F[x(\tau), \dot{x}(\tau), \tau] d\tau + E[x(t'+\delta), t'+\delta] - E[x(t'), t'] \right\} = 0 \quad (2-78)$$

Equation (2-78) is the discrete form of the dynamic programming condition for optimality and is frequently used in this form. The continuous form of this equation is derived by letting  $\delta$  approach zero. When this happens the terms in the braces are written

$$\int_{t'}^{t'+\delta} F[x(\tau), \dot{x}(\tau), \tau] d\tau = \delta \cdot F[x(t'), \dot{x}(t'), t'] + \delta^2 \cdot \{ \dots \} \quad (2-79)$$

$$E[x(t'+\delta), t'+\delta] - E[x(t'), t'] = \delta \cdot \frac{dE[x(t'), t']}{dt'} + \delta^2 \cdot \{ \dots \} \quad (2-80)$$

and the minimum performance criterion condition is

$$\delta \cdot \left\{ \min_{\dot{x}(t') \in S(t')} \left[ F[x(t'), \dot{x}(t'), t'] + \frac{dE[x(t'), t']}{dt'} + \delta \cdot \{ \dots \} \right] \right\} = 0 \quad (2-81)$$

Since  $\delta$  is arbitrary although small, this final condition, where  $\delta \rightarrow 0$ , is simply

$$\min_{\dot{x}(\tau) \in S(\tau)} \left\{ F[x(\tau), \dot{x}(\tau), \tau] + \frac{dE[x(\tau), \tau]}{d\tau} \right\} = 0 \quad (2-82)$$

where the variable  $t'$  has been replaced by  $\tau$ . Taking the total time derivative of the second term,

$$\frac{dE[x(\tau), \tau]}{d\tau} = \frac{\partial E[x(\tau), \tau]}{\partial \tau} + \dot{x}(\tau) \frac{\partial E[x(\tau), \tau]}{\partial x(\tau)} \quad (2-83)$$

the continuous dynamic programming necessary condition for a minimum performance criterion is

$$\min_{\dot{x}(\tau) \in S(\tau)} \left\{ F[x(\tau), \dot{x}(\tau), \tau] + \dot{x}(\tau) \frac{\partial E[x(\tau), \tau]}{\partial x(\tau)} \right\} = - \frac{\partial E[x(\tau), \tau]}{\partial \tau} \quad (2-84)$$

since  $\partial E/\partial \tau$  is not dependent upon  $\dot{x}(\tau)$ . The sufficient condition for a minimum is that the performance criterion is a continuous, strictly convex function of  $\dot{x}(\tau)$  and is represented

$$\frac{\partial^2}{\partial \dot{x}(\tau)^2} \left\{ F[x(\tau), \dot{x}(\tau), \tau] + \dot{x}(\tau) \frac{\partial E[x(\tau), \tau]}{\partial x(\tau)} \right\} = \frac{\partial^2 F[x(\tau), \dot{x}(\tau), \tau]}{\partial \dot{x}(\tau)^2} > 0 \quad (2-85)$$

The extension of the necessary condition of dynamic programming to the  $N^{\text{th}}$ -order dynamic process is here carried out in vector notation. The minimum performance criterion is defined as

$$E[\underline{x}(t'), t] = \min_{\underline{u}(\tau) \in U(\tau)} \int_{t'}^T H[\underline{x}(\tau), \underline{u}(\tau), \tau] d\tau \quad (2-86)$$

with the boundary condition  $E[\underline{x}(T), T] = 0$ . The continuous form of the minimum performance criterion is here seen to be equivalent to equation (2-82) and is

$$\min_{\underline{u}(\tau) \in U(\tau)} \left\{ H[\underline{x}(\tau), \underline{u}(\tau), \tau] + \frac{dE[x(\tau), \tau]}{d\tau} \right\} = 0 \quad (2-87)$$

The total time derivative equivalent of equation (2-82) is

$$\frac{dE[\underline{x}(\tau), \tau]}{d\tau} = \frac{\partial E[\underline{x}(\tau), \tau]}{\partial \tau} + \sum_{i=1}^N \dot{x}_i(\tau) \frac{\partial E[\underline{x}(\tau), \tau]}{x_i(\tau)} \quad (2-88)$$

and the condition for minimum performance which corresponds to equation (2-84) is

$$\begin{aligned} \min_{\underline{u}(\tau) \in U(\tau)} \left\{ H[\underline{x}(\tau), \underline{u}(\tau), \tau] + \sum_{i=1}^N f_i[\underline{x}(\tau), \underline{u}(\tau), \tau] \frac{\partial E[\underline{x}(\tau), \tau]}{\partial x_i(\tau)} \right\} \\ = - \frac{\partial E[\underline{x}(\tau), \tau]}{\partial \tau} \end{aligned} \quad (2-89)$$

In the dynamic programming formulation the term  $\partial E / \partial x_i$  is equivalent to the Lagrange multiplier,  $\lambda_i$ , of the calculus of variations and Pontryagin's maximum principle. From equation (2-89) it is not hard to understand why the discrete formulation is most frequently used to find the minimizing control for higher-than-first-order dynamic processes.

## Chapter 3

### DESIGN OF OPTIMAL FEEDBACK CONTROL SYSTEMS

#### Introduction

In this chapter the application of the optimal control methods of the previous chapter are applied to the problem of designing an optimal feedback controller. The general performance criterion for this work is an integral-square-error criterion weighting perturbations of both state and control variables from the nominal values of these variables. In various other applications of optimal control theory to nuclear reactor processes, the methods of the previous chapter were used to find the nominal state and control variables, but in this work these trajectories are assumed to have been previously determined.

The first consideration is the selection of the appropriate performance criterion, henceforth referred to as an error criterion. In this section is discussed such topics as control saturation constraints, selection of the weighting factors, and the incomplete measurement of the physical state variables.

Following is a section on the synthesis of the subclass of linear dynamic processes. The feedback controller as determined by either the maximum principle or dynamic programming are identical for linear systems, and only the

dynamic programming format is used. The general properties of linear optimal control systems are outlined specifically in this section. A discussion is also included on the stability of the linear optimal control system.

Finally, the synthesis of non-linear control systems is discussed. Here techniques for determining quasi-optimum feedback control systems are presented. Dynamic programming and the maximum principle are discussed separately for quasi-optimum feedback control because significant differences arise. Primarily, the maximum principle yields a quasi-optimum control which approximates the optimum control equation by a Taylor series expansion about the nominal trajectories. Dynamic programming, however, approximates the minimum error function (performance criterion) by a Taylor series expansion about the nominal trajectories. The differences in the two methods are not obvious a priori. Thus non-linear quasi-optimal feedback control synthesis should be evaluated by both of these methods to determine which is better for the problem at hand. Examples in chapter 4 demonstrate that the choice is not unique and that generalizations concerning system performance are somewhat difficult to assay beforehand.

### Feedback Control Design Considerations

The first consideration in the design of any control system is the selection of an acceptable performance criterion. For the feedback controller evaluation some measure of

the errors relative to the nominal variables of the dynamic process is required. Hence, a suitable performance criterion is an error criterion. In some problems this criterion might be specified only at a single point in time. For example it might be desired to minimize errors in the terminal values of the state variables. Such an error criterion would take the form

$$e(t) = \sum_{i=1}^N F_1 [x_i^*(T) - x_i(T)] \quad (3-1)$$

where the functions  $F_1$  are arbitrary, but would naturally consider only the magnitudes of the errors. Acceptable choices of  $F_1$  might be

$$F(y) = |y|; F(y) = y^2; F(y) = y^{2n} \quad (3-2)$$

An error measure that is equivalent to equation (3-1) is the impulse error measure

$$e(t) = \sum_{i=1}^N \int_t^T \delta(\tau-T) F_1 [x_i^*(\tau) - x_i(\tau)] d\tau \quad (3-3)$$

where  $\delta(\tau-T)$  is the unit-impulse function or the Dirac delta function. The upper limit is considered the time where control effort terminates.

A somewhat better error criterion, would be to minimize the accumulated errors for the entire control interval. Such a controller, in the true sense, is a feedback controller such that all  $x_i(t) \approx x_i^*(t)$  for all real time. Thus the errors are weighted over all future

time starting with the present. The error criterion for such a system is

$$e(t) = \sum_{i=1}^N \int_t^T \phi_i(\tau) F_i [x_i^*(\tau) - x_i(\tau)] d\tau \quad (3-4)$$

where the  $\phi_i(\tau)$  are weights assigned to allowable state variable errors. They are here termed state variable weighting factors.

In addition to minimizing state variable errors, it is usually desirable to minimize control variable errors also. It was stated in the section on the maximum principle that when control effort is not weighted in the performance criterion the optimal controller is a bang-bang controller. Such control is not always satisfactory, and when nominal control curves are available a servo-system is required such that  $u_j(t) \approx u_j^*(t)$  for all time.

The error measure of equation (3-4) is now modified to include control variable errors

$$e(t) = \sum_{i=1}^N \int_t^T \phi_i(\tau) F_i [x_i^*(\tau) - x_i(\tau)] d\tau + \sum_{j=1}^M \int_t^T \psi_j(\tau) G_j [u_j^*(\tau) - u_j(\tau)] d\tau \quad (3-5)$$

where the functions  $\psi_j(\tau)$  are termed the control variable weighting factors.

In this work quadratic-error criteria are used exclusively. There are several distinct advantages for

using such criteria. Firstly, large errors are penalized more heavily than smaller ones and this is desirable. The criterion can easily be chosen to be a strictly convex function of the errors, since the square terms are always positive definite, which is a basic condition for the error measure to assume a minimum value. Most importantly, however, the resulting optimal feedback controller is linear for linear dynamic processes and the resulting quasi-optimal controller using the maximum principle is linear for non-linear dynamic processes. A form of this quadratic-error criterion is

$$\begin{aligned}
 e(t) = & \sum_{i=1}^N \int_t^T \phi_{ii}(\tau) [x_i^*(\tau) - x_i(\tau)]^2 d\tau \\
 & + \sum_{j=1}^M \int_t^T \psi_{jj}(\tau) [u_j^*(\tau) - u_j(\tau)]^2 d\tau
 \end{aligned}
 \tag{3-6}$$

The conditions  $\phi_{ii}(\tau) \geq 0$  and  $\psi_{jj}(\tau) > 0$  are imposed so that the integrand is positive and strictly convex. In equation (3-6) cross-product terms between two different state variable errors, or between state variable and control variable errors, are not included, because they usually are meaningless in the design problem. Equation (3-6) is the basic error criterion used in all of the present examples.

Some dynamic processes require a large number of state variables for an accurate mathematical description. Unfortunately, the measurement of all of these variables may not be possible. For example in nuclear reactors



systems the direct measurement of delayed neutron precursor densities is not possible. These variables are required, however, for the feedback controller to be optimum. When incomplete measurement arises the most satisfactory means of obtaining these states is on-line computation of those not measured. This requires that at least one of the physical states be measured. Power level, for example, is a measurable state in the reactor and all precursor densities can be determined from this state.

Because of the frequent need to calculate the precursor concentrations in reactor kinetics, the on-line computer for these is evaluated here. The description of the simplest reactor with delayed neutrons (16) is

$$\dot{n} = \frac{\rho n}{\ell} - \frac{\beta n}{\ell} + \sum_{i=1}^6 \lambda_i c_i \quad (3-7)$$

$$\text{and} \quad \dot{c}_i = \frac{\beta_i n}{\ell} - \lambda_i c_i \quad i = 1, 2, \dots, 6 \quad (3-8)$$

where  $n$  is the reactor power, a state variable,  $\rho$  is the total reactivity, a control variable, and  $c_i$  are the precursor densities. The parameters  $\beta$ ,  $\beta_i$ ,  $\lambda_i$  and  $\ell$  are characteristic of the given reactor dynamic process and the type of fuel utilized.

$\beta_i$  = delayed neutron yield of  $i^{\text{th}}$  precursor-group per fission

$\beta = \sum_{i=1}^6 \beta_i$  = total delayed neutron precursor yield per fission

$\lambda_1$  = decay constant of the 1<sup>th</sup> precursor group

and  $\ell$  = the neutron lifetime of the reactor.

Equations (3-12) relate the state variables of precursor densities to the reactor power level and are linear differential equations. The transfer function of each group, with reactor power as an input, is

$$\frac{C_1(s)}{N(s)} = \frac{\beta_1/\ell\lambda_1}{s/\lambda_1 + 1} \quad i = 1, 2, \dots, 6 \quad (3-9)$$

The equivalent electrical analog with the same transfer impedance is represented in figure 1. The transfer function of the electrical network of figure 1 is

$$\frac{V_{C_1}(s)}{V_n(s)} = \frac{R_{21}/(R_{11} + R_{21})}{(R_{21}/(R_{11} + R_{21}))R_{11}C_1s + 1} \quad (3-10)$$

The output voltage  $V_{C_1}(t)$  corresponds to  $[c_1(t) - c_1(0)]$  when  $V_n(t)$  corresponds to  $[n(t) - n(0)]$  and the following equalities hold;

$$R_{21}/(R_{11} + R_{21}) = \beta_1/\ell\lambda_1 \quad (3-11)$$

$$R_{11}C_1 = \ell/\lambda_1 \quad (3-12)$$

Figure 2 is a schematic of the reactor plant with this type of on-line computer.

The selection of the weighting factors  $\phi_{1i}(\tau)$  and  $\psi_{jj}(\tau)$  can generally be specified by the performance

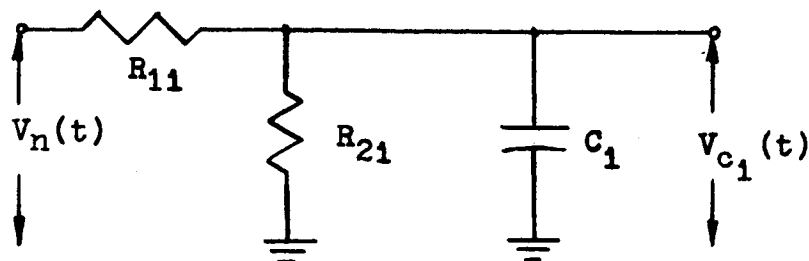


Figure 1

Electrical Network with Transfer Function Equivalent to Equation (3-9)

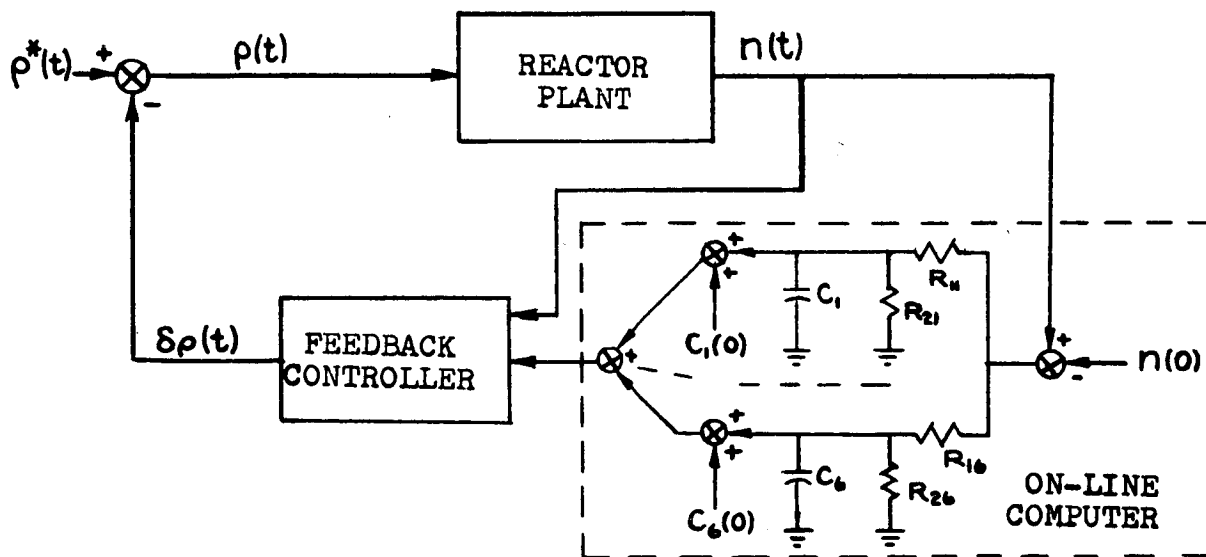


Figure 2

Schematic of Reactor with On-Line Computer

requirements, or constraints, of the overall dynamic process. The selection is not unique, and some alteration may be required of the initial choices. The more difficult the design problem in general, the more important the selection becomes if the design requirements are to be met.

A purely heuristic method of selecting the weighting factors is presented here in terms of elementary concepts. The instantaneous error measure is written

$$\begin{aligned}
 H(t) = & \sum_{i=1}^N \phi_{ii}(t) [x_i^*(t) - x_i(t)]^2 \\
 & + \sum_{j=1}^M \psi_{jj}(t) [u_j^*(t) - u_j(t)]^2
 \end{aligned} \tag{3-13}$$

The weighting factors are then determined on the basis of maximum allowable errors. For example, the maximum allowable state variable errors at any point in time contribute equally to the error measure, since it is desired to minimize the integrated sum. This is stated as

$$\phi_{ii}(\tau) = \left[ \frac{\delta x_N(\tau)_{MA}}{\delta x_i(\tau)_{MA}} \right]^2 \phi_{NN}(\tau) \quad i = 1, 2, \dots, N-1 \tag{3-14}$$

The  $\delta x_i(\tau)_{MA}$  are then

$$\delta x_i(\tau)_{MA} = [x_i^*(\tau) - x_i(\tau)] \text{ max allowable} \tag{3-15}$$

The same logic may be applied to the control errors

$$\psi_{jj}(\tau) = \left[ \frac{\delta u_1(\tau)_{MA}}{\delta u_j(\tau)_{MA}} \right]^2 \psi_{11}(\tau) \quad j = 2, 3, \dots, M \quad (3-16)$$

where  $\delta u_j(\tau)_{MA} = [u_j^*(\tau) - u_j(\tau)]$  max available (3-17)

The maximum allowable and available errors are subject to the design performance specifications.

In addition the total contribution due to maximum allowable state variable errors in the error measure should roughly equal the total contribution due to maximum available control errors in order to minimize the error criterion.

Using this relation

$$\sum_{i=1}^N \phi_{11}(\tau) \delta x_i^2(\tau)_{MA} = \sum_{j=1}^M \psi_{jj}(\tau) \delta u_j^2(\tau)_{MA} \quad (3-18)$$

When equations (3-14) and (3-16) are considered equation (3-18) is equivalent to the following relation

$$\phi_{NN}(\tau) = \frac{M}{N} \left[ \frac{\delta u_1(\tau)_{MA}}{\delta x_N(\tau)_{MA}} \right]^2 \psi_{11}(\tau) \quad (3-19)$$

Finally, if  $\psi_{11}(\tau) = 1$ , all other weighting factors are uniquely specified by this procedure. Probably, however, the maximum errors do not all occur simultaneously and the equations are somewhat unrealistic. The equations indicate that the weighting factors are time-varying, but they can be treated as constants when  $\delta x_i(\tau)_{MA}$  and  $\delta u_j(\tau)_{MA}$  are set equal to their largest values during the control process.

In some instances it is desirable to weight the terminal state variable errors more heavily than at any other time in the control interval. This is accomplished by impulse weighting of the type considered in equation (3-3). If this is necessary, the weighting functions for the state variables become

$$\phi'_{i1}(\tau) = \phi_{i1}(\tau) + \phi_{i,T} \delta(\tau-T) \quad (3-20)$$

where  $\phi_{i,T}$  is the impulse weighting factor for the  $i^{\text{th}}$ -state variable error at the terminal time. Impulse weighting has the same effect as increasing the duration of the control interval.

In the design of feedback control systems, control and state variable saturation constraints must not be violated. The error criterion of equation (3-6) does not take into consideration saturation constraints which are sometimes termed "hard" constraints. A simple procedure is available (7) to incorporate these constraints into the error criterion of equation (3-6). This is a technique whereby hard constraints are replaced by mathematical relations, termed "soft" constraints, that heavily penalize values near the limits of saturation. However, in this work soft constraints have not been considered.

When reactivity is the control variable in a nuclear reactor, the resulting dynamic process, as indicated by equation (3-7), is non-linear. This might lead one to

believe that only non-linear synthesis should be considered for nuclear reactor processes. However, Kliger (5) has introduced a technique which transforms the non-linear dynamic process into a linear one very simply. The non-linear dynamic process can be written alternately as

$$\dot{\underline{x}}(t) = A(t)\underline{x}(t) + \underline{b}[\underline{x}(t), \underline{u}(t), t] \quad (3-21)$$

where the first term on the right represents all the linear terms in the state vector  $\underline{x}(t)$  and the vector  $\underline{b}$  represents the nonlinearities and the control terms of the dynamic process. A pseudo-control vector is then equated to the vector  $\underline{b}$

$$\underline{u}'(t) = \underline{b}[\underline{x}(t), \underline{u}(t), t] \quad (3-22)$$

and the resulting dynamic process is linear.

For a nuclear reactor described by the equations

$$\dot{n} = \frac{\rho n}{\ell} - \frac{\beta n}{\ell} + \lambda c \quad (3-23)$$

$$\dot{c} = \frac{\beta n}{\ell} - \lambda c \quad (3-24)$$

the pseudo-control variable would be

$$u'(t) = \frac{\rho(t)n(t)}{\ell} \quad (3-25)$$

Since  $n(t)$  is a measurable state, the actual control reactivity is obtained by multiplying  $u'(t)$  by the measurable

quantity  $\ell/n(t)$ . This necessarily requires that a small auxiliary component be introduced into the control system to carry out this operation. Figure 3 is a schematic of the control system for equations (3-23) and (3-24).

### Synthesis of Linear Systems

In this section the optimal feedback controller equations are developed for dynamic processes which are described by equation (2-5), here repeated

$$\dot{\underline{x}}(t) = A(t)\underline{x}(t) + B(t)\underline{u}(t) \quad (3-26)$$

where  $A(t)$  and  $B(t)$  are given by equation (2-6). The variational problem is to minimize the error criterion of equation (3-6) subject to the linear dynamic process of equation (3-26). The dynamic programming format of equation (2-89) is used here. The error criterion is repeated here

$$\begin{aligned} e(t) = & \sum_{i=1}^N \int_t^T \phi_{11}(\tau) [x_1^*(\tau) - x_1(\tau)]^2 d\tau \\ & + \sum_{j=1}^M \int_t^T \psi_{jj}(\tau) [u_j^*(\tau) - u_j(\tau)]^2 d\tau \end{aligned} \quad (3-27)$$

The development of the optimum controller equations can be presented in a concise way using vector-matrix notation.

Equation (3-27) rewritten in vector-matrix form is

$$e(t) = \int_t^T \left\{ [\underline{x}^* - \underline{x}]^T \Phi [\underline{x}^* - \underline{x}] + [\underline{u}^* - \underline{u}]^T \Psi [\underline{u}^* - \underline{u}] \right\} d\tau \quad (3-28)$$



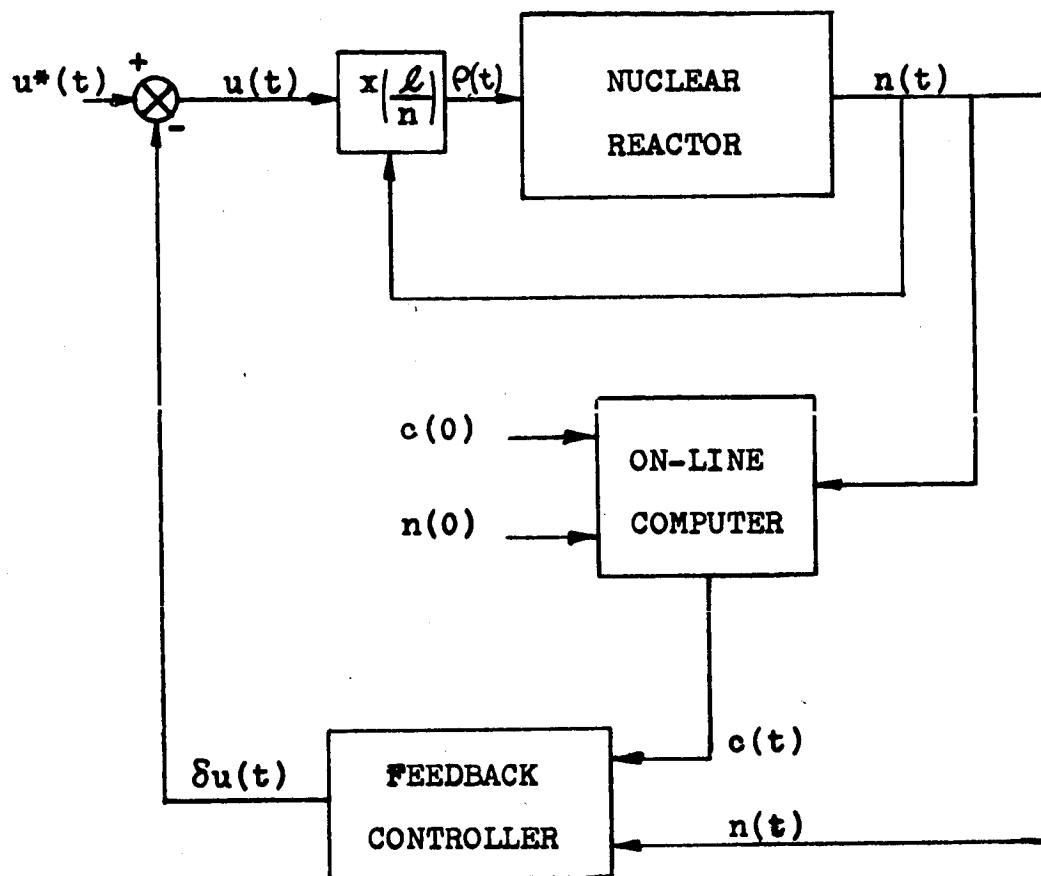


Figure 3

Reactor System with a Pseudo-Control Variable

where the time arguments have been dropped for conciseness. The superscript T indicates matrix transpose. The weighting factor matrices  $\Phi$  and  $\Psi$  are diagonal of form

$$\Phi = \begin{bmatrix} \phi_{11}(\tau) & 0 & \dots & 0 \\ 0 & \phi_{22}(\tau) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_{NN}(\tau) \end{bmatrix} \quad (3-29)$$

$$\text{and } \Psi = \begin{bmatrix} \psi_{11}(\tau) & 0 & \dots & 0 \\ 0 & \psi_{22}(\tau) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_{MM}(\tau) \end{bmatrix} \quad (3-30)$$

Using the dynamic programming format of equation (2-89) in vector-matrix form, the necessary condition for a minimum-error controller is

$$\min_{\underline{u}(\tau)} \left\{ [\underline{x}^* - \underline{x}]^T \Phi [\underline{x}^* - \underline{x}] + [\underline{u}^* - \underline{u}]^T \Psi [\underline{u}^* - \underline{u}] + \dot{\underline{x}}^T \left[ \frac{\partial E}{\partial \underline{x}} \right] \right\} = - \left[ \frac{\partial E}{\partial \tau} \right] \quad (3-31)$$

$$\text{where } \left[ \frac{\partial E}{\partial \underline{x}} \right]^T = \left[ \frac{\partial E}{\partial x_1} \quad \frac{\partial E}{\partial x_2} \quad \dots \quad \frac{\partial E}{\partial x_N} \right] \quad (3-32)$$

The optimal control vector is determined when the partial derivative with respect to  $\underline{u}(t)$  of the term within the braces of equation (3-31) is taken

$$\frac{\partial}{\partial \underline{u}} \left\{ \left[ \underline{u}^* - \underline{u} \right]^T \Psi \left[ \underline{u}^* - \underline{u} \right] + \dot{\underline{x}}^T \left[ \frac{\partial E}{\partial \underline{x}} \right] \right\} = 0 \quad (3-33)$$

$$\text{where } \dot{\underline{x}}^T = \underline{x}^T A^T + \underline{u}^T B^T \quad (3-34)$$

Performing this operation the optimal control vector is

$$\underline{u}^0(\tau) = \underline{u}^* - \frac{1}{2} \Psi^{-1} B^T \left[ \frac{\partial E}{\partial \underline{x}} \right] \quad (3-35)$$

where the superscript  $^0$  designates the minimizing control. Substituting equation (3-35) into (3-31) the condition for a minimum is obtained.

$$\begin{aligned} & \left[ \underline{x}^* - \underline{x} \right]^T \Phi \left[ \underline{x}^* - \underline{x} \right] - \frac{1}{4} \left[ \frac{\partial E}{\partial \underline{x}} \right]^T B \Psi^{-1} B^T \left[ \frac{\partial E}{\partial \underline{x}} \right] \\ & + \underline{x}^T A^T \left[ \frac{\partial E}{\partial \underline{x}} \right] + \underline{u}^{*T} B^T \left[ \frac{\partial E}{\partial \underline{x}} \right] = - \frac{\partial E}{\partial \tau} \end{aligned} \quad (3-36)$$

The problem now is reduced to finding  $E[\underline{x}, \tau]$ . The solution is obtained by assuming that the mathematical form of  $E[\underline{x}, \tau]$  is a quadratic function of  $\underline{x}(\tau)$  with time dependent coefficients. The assumed solution is then substituted into equation (3-36) and shown to be the correct solution under certain conditions.

The assumed solution is

$$E[\underline{x}(\tau), \tau] = k(\tau) - \underline{k}^T(\tau) \underline{x} - \underline{x}^T \underline{k}(\tau) + \underline{x}^T \underline{k}(\tau) \underline{x} \quad (3-37)$$

where  $k(\tau)$  is a scalar function, and

$$\underline{k}^T(\tau) = [k_1(\tau) k_2(\tau) \dots k_N(\tau)] \quad (3-38)$$

and

$$K(\tau) = K^T(\tau) = \begin{bmatrix} k_{11}(\tau) & k_{12}(\tau) & \dots & k_{1N}(\tau) \\ k_{12}(\tau) & k_{22}(\tau) & \dots & k_{2N}(\tau) \\ \vdots & \vdots & \ddots & \vdots \\ k_{1N}(\tau) & k_{2N}(\tau) & \dots & k_{NN}(\tau) \end{bmatrix} \quad (3-39)$$

From equation (3-37) the necessary partial derivatives of equation (3-36) are

$$\frac{\partial E}{\partial \underline{x}} = -2\underline{k}(\tau) + 2\underline{k}\underline{x} \quad (3-40)$$

and

$$\frac{\partial E}{\partial \tau} = \dot{\underline{k}}(\tau) - \underline{\dot{k}}^T(\tau)\underline{x} - \underline{x}^T\underline{\dot{k}}(\tau) + \underline{x}^T\dot{K}(\tau)\underline{x} \quad (3-41)$$

Substitution of equation (3-40) and (3-41) into (3-36) gives the necessary condition in terms of  $\underline{k}(\tau)$ ,  $\underline{\dot{k}}(\tau)$  and  $K(\tau)$ . The result is

$$\begin{aligned} & \underline{x}^{*T}\underline{\Phi}\underline{x}^* - \underline{x}^{*T}\underline{\Phi}\underline{x} - \underline{x}^T\underline{\Phi}\underline{x}^* + \underline{x}^T\underline{\Phi}\underline{x} - \underline{x}^T\underline{K}\underline{B}\Psi^{-1}\underline{B}^T\underline{K}\underline{x} + \underline{k}^T\underline{B}\Psi^{-1}\underline{B}^T\underline{k}\underline{x} \\ & + \underline{x}^T\underline{K}\underline{B}\Psi^{-1}\underline{B}^T\underline{k} - \underline{k}^T\underline{B}\Psi^{-1}\underline{B}^T\underline{k} + \underline{x}^T\underline{K}\underline{A}\underline{x} - \underline{k}\underline{A}\underline{x} + \underline{x}^T\underline{K}\underline{B}\underline{u}^* - \underline{k}^T\underline{B}\underline{u}^* \\ & + \underline{x}^T\underline{A}^T\underline{K}\underline{x} + \underline{u}^{*T}\underline{K}\underline{x} - \underline{x}^T\underline{A}^T\underline{k} - \underline{u}^{*T}\underline{B}^T\underline{k} = -\dot{\underline{k}} + \underline{x}^T\dot{\underline{k}} + \dot{\underline{k}}^T\underline{x} - \underline{x}^T\dot{K}\underline{x} \end{aligned} \quad (3-42)$$

The vector  $\underline{x}(\tau)$  can assume any arbitrary value and thus, in order for equation (3-42) to represent a minimum, the coefficients of the powers of  $\underline{x}(\tau)$  on the left must equal the coefficients on the right. These conditions result in first-order differential equations which describe the  $\underline{k}$  parameters.

$$-\dot{\underline{k}} = \underline{x}^{*T} \Phi \underline{x}^* - \underline{k}^T B \Psi^{-1} B^T \underline{k} - 2 \underline{k}^T B \underline{u}^* \quad (3-43)$$

$$-\dot{\underline{k}} = \Phi \underline{x}^* - KB \Psi^{-1} B^T \underline{k} + A^T \underline{k} + KB \underline{u}^* \quad (3-44)$$

and 
$$-\dot{\underline{k}} = KA + A^T \underline{k} - KB \Psi^{-1} B^T \underline{k} + \Phi \quad (3-45)$$

The boundary conditions for the equations are found from the boundary condition for the minimum error criterion. The error functional  $E$  is given by equation (2-86) such that

$$E[\underline{x}(\tau'), \tau'] = \min_{\underline{u}(\tau)} \int_{\tau'}^T H[\underline{x}, \underline{u}, \tau] d\tau \quad (3-46)$$

and from equation (3-28) the identical relation is

$$E[\underline{x}(\tau'), \tau'] = \min_{\underline{u}(\tau)} e(\tau') \quad (3-47)$$

Thus the boundary conditions are determined from the condition

$$E[\underline{x}(T), T] = 0 \quad (3-48)$$

or, if equation (3-20) is utilized as the state variable weighting function, impulse weighting of the terminal error implies

$$E[\underline{x}(T), T] = [\underline{x}(T) - \underline{x}^*(T)]^T \Phi_T [\underline{x}(T) - \underline{x}^*(T)] \quad (3-49)$$

where

$$\Phi_T = \begin{bmatrix} \phi_{1,T} & 0 & \dots & 0 \\ 0 & \phi_{2,T} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_{N,T} \end{bmatrix} \quad (3-50)$$

From equation (3-37) the boundary conditions of equations (3-43), (3-44) and (3-45) are seen to be

$$\underline{k}(T) = \underline{x}^*(T) \Phi_T \underline{x}^*(T) \quad (3-51)$$

$$\underline{k}(T) = \Phi_T \underline{x}^*(T) \quad (3-52)$$

and

$$K(T) = \Phi_T \quad (3-53)$$

or they are all zero if equation (3-48) holds. It is easy to demonstrate that equations (3-43), (3-44) and (3-45) imply the following relations:

$$\underline{k}(\tau) = K(\tau) \underline{x}^*(\tau) \quad (3-54)$$

and

$$\underline{k}(\tau) = \underline{x}^{*T}(\tau) K(\tau) \underline{x}^*(\tau) \quad (3-55)$$

and hence only the solution of equation (3-45) is required, and is here repeated

$$-\dot{K} = KA + A^T K - KB\Psi^{-1}B^T K + \Phi \quad (3-56)$$

where

$$K(T) = \Phi_T.$$

Equation (3-56) is termed a matrix Riccati differential equation. The solution of these simultaneous differential equations for greater than a second-order

dynamic process requires rather sophisticated digital computer programs or analog computer solutions. For time-invariant dynamic processes, the matrices A and B have constant elements, and equation (3-56) is non-linear with constant coefficients.

Equation (3-32) gives the optimal control vector. Utilizing equations (3-40) and (3-54) the optimal control is expressed as

$$\underline{u}(\tau) = \underline{u}^*(\tau) + \Psi^{-1} B^T K(\tau) [\underline{x}^*(\tau) - \underline{x}(\tau)] \quad (3-57)$$

Thus the optimal controller consists of the nominal control vector, plus a feedback element that weights linear perturbations in the state variables from their nominal values. The term  $\Psi^{-1} B^T K(\tau)$  is the optimal feedback gain and is seen to be time-varying. Figure 4 is a schematic of the optimally controlled linear dynamic process.

It is worth noting here that the solution of  $K(\tau)$  does not require prior knowledge of the nominal state and control vectors. An equation similar to equation (3-56) must be solved when the approximation technique of Pontryagin's maximum principle is applied to non-linear system synthesis. However, the matrices  $A(t)$  and  $B(t)$  are time-varying functions determined by  $\underline{x}^*(t)$  and  $\underline{u}^*(t)$  and are different than the corresponding matrices for equation (3-56). The main advantage of the linear optimal feedback controller then is that, once equation (3-56) is evaluated

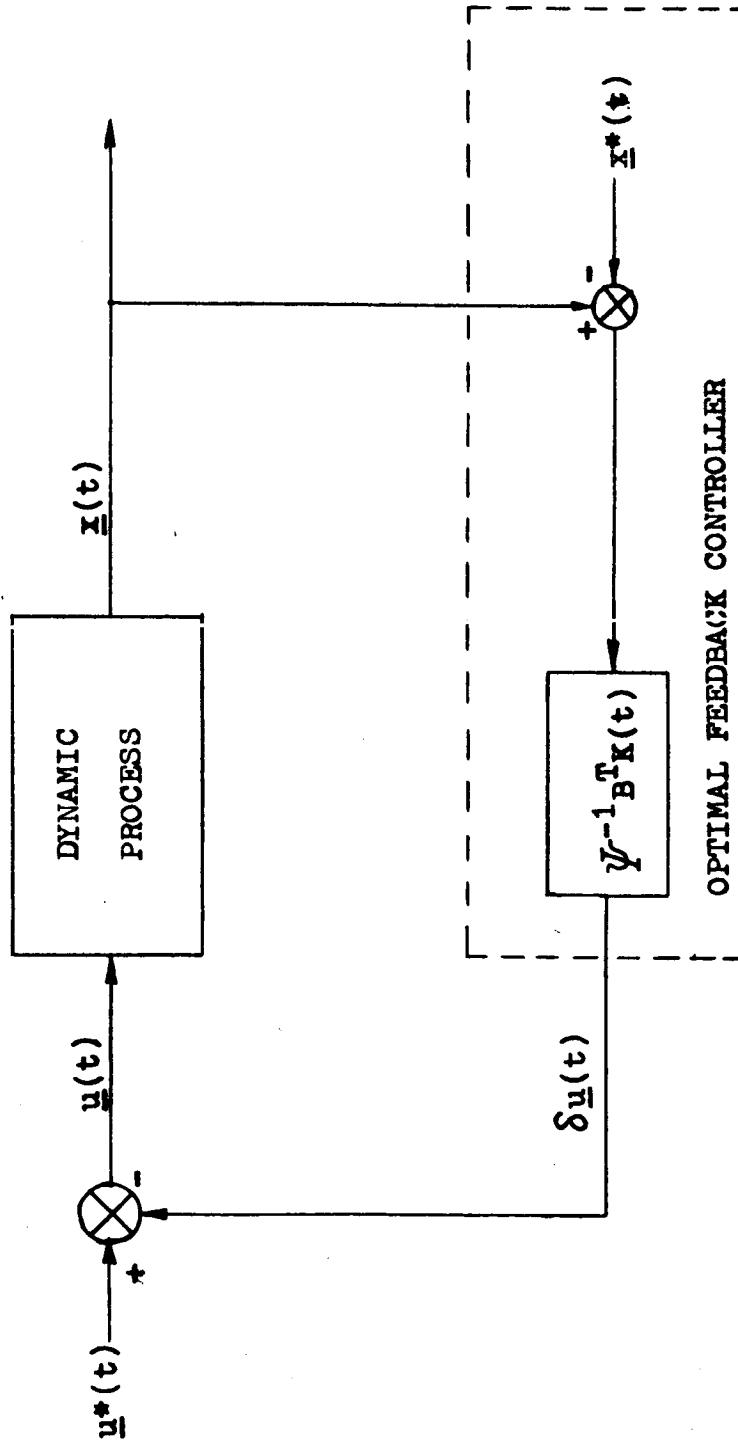


Figure 4

Optimal Feedback Controller



and the gains are determined, this controller holds for any nominal set of values the dynamic process may undergo.

The question of stability of the control system often arises. It can be easily shown that a linear optimal control system based on the error criterion of equation (3-28) is asymptotically stable in the large for certain feedback conditions. The system will be considered stable if  $\underline{x}(t) \rightarrow \underline{x}^*(t)$  as  $t \rightarrow \infty$ . A new equilibrium vector representing the state variable error is defined

$$\underline{z}(\tau) = \underline{x}(\tau) - \underline{x}^*(\tau) \quad (3-58)$$

Utilizing equations (3-34) and (3-57) the newly defined dynamic process is

$$\begin{aligned} \dot{\underline{z}}(\tau) &= \dot{\underline{x}}(\tau) - \dot{\underline{x}}^*(\tau) \\ &= \left\{ \underline{A}\underline{x}(\tau) + \underline{B}\underline{u}^*(\tau) - \underline{B}\Psi^{-1}\underline{B}^T\underline{K}(\tau) [\underline{x}(\tau) - \underline{x}^*(\tau)] \right\} \\ &\quad - \left\{ \underline{A}\underline{x}^*(\tau) + \underline{B}\underline{u}^*(\tau) \right\} \\ &= \left[ \underline{A} - \underline{B}\Psi^{-1}\underline{B}^T\underline{K}(\tau) \right] \underline{z}(\tau) \end{aligned} \quad (3-59)$$

The solution of this equation is

$$\underline{z}(t) = \exp \int_0^t \left[ \underline{A} - \underline{B}\Psi^{-1}\underline{B}^T\underline{K}(\tau) \right] d\tau \underline{z}(0) \quad (3-60)$$

or in eigenvalue notation

$$\underline{z}(t) = \exp \int_0^t \theta(\tau) d\tau \underline{z}(0) \quad (3-61)$$

where  $\theta(\tau)$  is the diagonal matrix

$$\theta(\tau) = \begin{bmatrix} \theta_1(\tau) & 0 & \dots & 0 \\ 0 & \theta_2(\tau) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_N(\tau) \end{bmatrix} \quad (3-62)$$

The elements  $\theta_1(\tau)$  are found from

$$\left| \theta(\tau) - A + B\Psi^{-1}B^TK(\tau) \right| = 0 \quad (3-63)$$

and for stability all  $\theta_1(\tau)$  must be less than or equal to zero as  $t \rightarrow \infty$ . Where the system matrix  $A$  corresponds to a stable linear dynamic process this condition is automatically satisfied, since  $K(\tau)$  is determined for a strictly convex error criterion.

The Riccati matrix equation (3-56) reaches its steady state value as  $t \rightarrow \infty$ , and  $\dot{K} \rightarrow 0$ , and  $K(\tau)$  is a matrix of constant elements. When  $\dot{K} = 0$  the determinant of equation (3-63) can be rewritten.

$$\left| \theta + K^{-1}\dot{\Phi} + K^{-1}A^TK \right| = 0 \quad \text{as } t \rightarrow \infty \quad (3-64)$$

Thus, if the system matrix  $A$  corresponds to an unstable system without feedback, stability can be regained by proper selection of the matrix  $\dot{\Phi}$ . Due to the dependence of  $K$  on  $\dot{\Phi}$  and  $A$ , this selection is not obvious a priori.

It is possible for unstable control systems to result if the terminal control time  $T$  is too short.

Asymptotic stability is only guaranteed where control effort is applied until all  $\theta_1(t)$  become and remain negative semi-definite. For this reason the terminal time  $T$  should be chosen long enough that the feedback gains of equation (3-56) have begun to settle at their steady state values. If  $T$  is sufficiently long that the gains approximate their steady state values during most of the control interval, the steady state solution of equation (3-56) is sufficient. In this case the feedback gains are constant with respect to time which is desirable.

#### Synthesis of Non-Linear Quasi-Optimal Control Systems

If either the dynamic process is non-linear or the performance criterion is non-quadratic the preceding development does not hold. Closed-form expressions for the optimum control equation in non-linear problems cannot be analytically determined except for some simple examples. Therefore, approximation techniques must be used. In this section the control system is assumed to operate in a small region about the nominal state and control vectors  $\underline{x}^*(t)$  and  $\underline{u}^*(t)$ . This immediately sets the requirement that these vectors must be known beforehand to design the quasi-optimal feedback controller.

The synthesis problem is then to find a suitable approximation of the optimal control vector

$$\underline{u}^0(t) = \underline{P}[\underline{x}^*(t), t] \quad (3-65)$$

in terms of the measurable state vector  $\underline{x}(t)$ . Two methods of obtaining the approximation are currently used. The most obvious is a Taylor expansion of the optimal control vector about  $\underline{x}^*(t)$  confining the system to operate in a suitably small region. This is represented by

$$\underline{u}(t) \approx \underline{P}[\underline{x}^*(t), t] + \frac{\partial \underline{P}[\underline{x}^*(t), t]}{\partial \underline{x}^*(t)} [\underline{x}(t) - \underline{x}^*(t)] \quad (3-66)$$

where

$$\frac{\partial \underline{P}}{\partial \underline{x}^*} = \begin{bmatrix} \frac{\partial P_1}{\partial x_1} & \frac{\partial P_1}{\partial x_2} & \dots & \frac{\partial P_1}{\partial x_N} \\ \frac{\partial P_2}{\partial x_1} & \frac{\partial P_2}{\partial x_2} & \dots & \frac{\partial P_2}{\partial x_N} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \frac{\partial P_M}{\partial x_1} & \frac{\partial P_M}{\partial x_2} & \dots & \frac{\partial P_M}{\partial x_N} \end{bmatrix} \quad (3-67)$$

Another method is the use of a similar approximation for the minimum error function,  $E[\underline{x}(t), t]$ , instead of the optimum control equation.

These two methods are developed here using the maximum principle approach for the first and the dynamic programming principle for the latter. Later it will be shown that it is not easy to decide which quasi-optimal control yields the better controller unless each has been evaluated separately.

First, the maximum principle approach is considered. The dynamic process is represented by the set of non-linear

first-order differential equations

$$\dot{\underline{x}}(t) = \underline{f}[\underline{x}(t), \underline{u}(t), t] \quad (3-68)$$

It is desired that the system operate in a small region about the nominal trajectories and the perturbations are represented

$$\delta \underline{u}(t) = \underline{u}(t) - \underline{u}^*(t) \text{ and } \delta \underline{x}(t) = \underline{x}(t) - \underline{x}^*(t) \quad (3-64)$$

The resulting piecewise linear differential equations describing perturbations from the nominal variables are time-varying of the form

$$\begin{aligned} \delta \dot{\underline{x}}(t) = & \frac{\partial \underline{f}[\underline{x}^*(t), \underline{u}^*(t), t]}{\partial \underline{x}^*(t)} \delta \underline{x}(t) \\ & + \frac{\partial \underline{f}[\underline{x}^*(t), \underline{u}^*(t), t]}{\partial \underline{u}^*(t)} \delta \underline{u}(t) \end{aligned} \quad (3-70)$$

or 
$$\delta \dot{\underline{x}}(t) = A(t) \delta \underline{x}(t) + B(t) \delta \underline{u}(t) \quad (3-71)$$

where  $A(t)$  is the system matrix and  $B(t)$  is the input matrix. Both are now direct functions of  $\underline{x}^*(t)$  and  $\underline{u}^*(t)$ .

The quadratic performance index is written

$$e(t) = \frac{1}{2} \int_t^T \left\{ \delta \underline{x}^T \Phi \delta \underline{x} + \delta \underline{u}^T \Psi \delta \underline{u} \right\} d\tau \quad (3-72)$$

and the system is now subject to the constraints of equation (3-70). The Hamiltonian function then is

$$H^*(\tau) = \frac{1}{2} \left\{ \delta \underline{x}^T \Phi \delta \underline{x} + \delta \underline{u}^T \Psi \delta \underline{u} \right\} + \underline{\lambda}^T [A(\tau) \delta \underline{x} + B(\tau) \delta \underline{u}] \quad (3-73)$$

The necessary conditions of chapter 2 are applied to find the optimal control equation for  $\delta \underline{u}$ , and the adjoint variables  $\underline{\lambda}(t)$ . These equations are

$$\frac{\partial H'}{\partial \delta \underline{u}} = \Psi \delta \underline{u} + B^T \underline{\lambda} = 0 \quad (3-74)$$

$$\frac{\partial H'}{\partial \delta \underline{x}} = \Phi \delta \underline{x} + A^T \underline{\lambda} = -\dot{\underline{\lambda}} \quad (3-75)$$

The optimal control equation becomes

$$\delta \underline{u} = -\Psi^{-1} B^T \underline{\lambda} \quad (3-76)$$

The adjoint variables are assumed to have the form

$$\underline{\lambda}(t) = K(t) \delta \underline{x}(t) \quad (3-77)$$

where  $K(t)$  is a matrix of time-varying elements and can be found from equation (3-75)

$$-\dot{K}(t) \delta \underline{x}(t) - K \delta \dot{\underline{x}}(t) = \Phi \delta \underline{x} + A^T K \delta \underline{x} \quad (3-78)$$

The optimal control equation (3-76) is now written

$$\delta \underline{u} = -\Psi^{-1} B^T K \delta \underline{x} \quad (3-79)$$

Equation (3-71) describing the dynamic process becomes

$$\delta \dot{\underline{x}} = [A - B \Psi^{-1} B^T K] \delta \underline{x} \quad (3-80)$$

Substituting equation (3-80) into (3-78)

$$-\left[\dot{K} + KA - KB \Psi^{-1} B^T K\right] \delta \underline{x} = \left[\Phi + A^T K\right] \delta \underline{x} \quad (3-81)$$

Since  $\delta \underline{x}$  is arbitrary the above equation is rewritten

$$-\dot{K} = KA + A^T K - KB\Psi^{-1}B^T K + \Phi \quad (3-82)$$

This is seen to be the Riccati matrix equation (3-56) of the previous section. The difference, however, is that the matrices  $A(t)$  and  $B(t)$  are here explicit functions of the nominal vectors  $\underline{x}^*(t)$  and  $\underline{u}^*(t)$ . The optimal feedback control of equation (3-79) demonstrates that the feedback gain,  $-\Psi^{-1}B^T K$ , is now a function of the nominal variables also. Figure 4 of the previous section is the schematic for the controlled dynamic process of this approximate method.

From the transversality condition at the terminal time,  $\underline{\lambda}(T) = 0$ , and since  $\delta \underline{x}(T)$  is arbitrary the boundary condition for equation (3-82) is  $K(T) = 0$ , or  $K(T) = \Phi_T$  if impulse weighting of the terminal errors is required.

The second method of quasi-optimal control is developed using the dynamic programming format. The dynamic programming equation for a minimization for the non-linear problem is

$$\frac{\partial E[\underline{x}, \tau]}{\partial \tau} + \min_{\underline{u}(\tau) \in U} \left\{ H[\underline{x}, \underline{u}, \tau] + \underline{f}^T \frac{\partial E[\underline{x}, \tau]}{\partial \underline{x}} \right\} = 0 \quad (3-83)$$

where the Hamiltonian is defined

$$H' = H[\underline{x}, \underline{u}, \tau] + \underline{f}^T \frac{\partial E[\underline{x}, \tau]}{\partial \underline{x}} \quad (3-84)$$

The minimum value of the Hamiltonian is represented

$$H^{**} = \min_{\underline{u}(\tau) \in U} \left\{ H'[\underline{x}, \underline{u}, \tau] \right\} \quad (3-85)$$

and equation (3-83) is written as

$$\frac{\partial E[\underline{x}, \tau]}{\partial \tau} + H^{**}[\underline{x}, \tau] = 0 \quad (3-86)$$

The approximation procedure is now applied to equation (3-86) by making a Taylor series expansion of both the minimum error function and the Hamiltonian function about  $\underline{x}^*(t)$ . The results of this approximation procedure are presented here without the complete derivations (?).

An arbitrary  $p^{\text{th}}$ -degree expansion of the minimum error function is

$$\begin{aligned} E_p[\underline{x}, \tau] = & k - 2 \sum_{n_1=1}^N k_{n_1} x_{n_1} + \sum_{n_1=1}^N \sum_{n_2=1}^N k_{n_1 n_2} x_{n_1} x_{n_2} \\ & + \dots + \frac{2}{p} \sum_{n_1=1}^N \dots \sum_{n_p=1}^N k_{n_1 \dots n_p} x_{n_1} \dots x_{n_p} \end{aligned} \quad (3-87)$$

The minimum  $p^{\text{th}}$ -degree Hamiltonian is represented

$$H_p^{**}[\underline{x}, \tau] = \min_{\underline{u}(\tau) \in U} \left\{ H[\underline{x}, \underline{u}, \tau] + \underline{f}^T[\underline{x}, \underline{u}, \tau] \frac{\partial E_p[\underline{x}, \tau]}{\partial \underline{x}} \right\} = 0 \quad (3-88)$$

The approximate form of equation (3-86) then is

$$\frac{\partial E_p[\underline{x}, \tau]}{\partial \tau} + H_p^{**} + \left\{ (p+1)^{\text{st}}\text{-degree terms in } \delta \underline{x} \right\} = 0 \quad (3-89)$$



The power series expansion of  $H_p^{i*}$  can be written

$$\begin{aligned}
 H_p^{i*} = H_0^* &+ \sum_{n_1=1}^N H_{n_1}^* x_{n_1} + \sum_{n_1=1}^N \sum_{n_2=1}^N H_{n_1 n_2}^* x_{n_1} x_{n_2} \\
 &+ \dots + \sum_{n_1=1}^N \dots \sum_{n_p=1}^N H_{n_1 \dots n_p}^* x_{n_1} \dots x_{n_p}
 \end{aligned}
 \tag{3-90}$$

where the  $H^*$  functions are

$$\begin{aligned}
 H_{n_1 \dots n_j}^* &= \frac{1}{j!} \left\{ \frac{\partial^j H_p^{i*}}{\partial x_{n_1} \dots \partial x_{n_j}} - \sum_{n_{j+1}=1}^N \frac{\partial^{j+1} H_p^{i*}}{\partial x_{n_1} \dots \partial x_{n_{j+1}}} x_{n_{j+1}}^* \right. \\
 &+ \dots + \left. \frac{(-1)^{p-j}}{(p-j)!} \sum_{n_{j+1}=1}^N \dots \sum_{n_p=1}^N \frac{\partial^{p-j} H_p^{i*}}{\partial x_{n_1} \dots \partial x_{n_p}} x_{n_{j+1}}^* \dots x_{n_p}^* \right\} \Big|_{\underline{x}=\underline{x}^*}
 \end{aligned}
 \tag{3-91}$$

The function  $H_0^*$  is computed when  $j = 0$  in equation (3-91).

Substituting equations (3-90) and (3-87) into equation (3-89), the result, when the common coefficients of  $\underline{x}$  are collected, is

$$\begin{aligned}
 \left[ \dot{k} + H_0^* \right] - 2 \sum_{n_1=1}^N \left[ \dot{k}_{n_1} - \frac{1}{2} H_{n_1}^* \right] x_{n_1} + \dots \\
 + \frac{2}{p} \sum_{n_1=1}^N \dots \sum_{n_p=1}^N \left[ \dot{k}_{n_1 \dots n_p} + \frac{p}{2} H_{n_1 \dots n_p}^* \right] x_{n_1} \dots x_{n_p} = 0
 \end{aligned}
 \tag{3-92}$$

Finally, equation (3-92) is valid for all  $\underline{x}$  when

$$\dot{k} = - H_0^*
 \tag{3-93}$$

$$\dot{k}_{n_1} = \frac{1}{2} H_{n_1}^* \quad (3-94)$$

$$\text{and } \dot{k}_{n_1 \dots n_j} = -\frac{p}{2} H_{n_1 \dots n_j}^* \quad j = 2, 3, \dots, p \quad (3-95)$$

The boundary conditions on the above equation are

$$E_p [\underline{x}(T), T] = [\underline{x}(T) - x^*(T)]^T \dot{\Phi}_T [x(T) - x^*(T)] \quad (3-96)$$

When the operations of this method are carried out, equations (3-93), (3-94) and (3-95) are generally non-linear, containing time-varying functions of  $\underline{x}^*(\tau)$  and  $\underline{u}^*(\tau)$ .

Except for the computational difficulties of solving the  $k$  parameters, this method is a very flexible synthesis technique for quasi-optimal non-linear control. Where non-quadratic performance criteria are used, this technique is superior to that of the maximum principle.

## Chapter 4

### FIRST-ORDER NUCLEAR REACTOR SYNTHESIS

#### Introduction

In this chapter several first-order reactor dynamic processes are considered, to introduce and demonstrate the application of optimal feedback control. The considerations of the previous chapter are applied to three separate first-order mathematical reactor models. Both linear and non-linear synthesis of these models are evaluated. The problem is to determine the mathematical form of the compensating control reactivity which minimizes the integrated errors of reactor power and control reactivity.

The nominal power for all cases is chosen to be a constant steady state value of  $n^*(t) = n_0$ . This is necessary because usually first order models do not adequately represent reactor dynamics during transient operation. The majority of the work is based on an infinite control interval, i.e.  $T = \infty$ . Saturation constraints are also considered.

For each model a comparison is made between the control reactivity determined by linear synthesis, exact non-linear synthesis, and the quasi-optimal methods of non-linear synthesis discussed in the previous chapter.

### Linear Prompt Neutron Reactor Synthesis

The model chosen here to represent the nuclear reactor dynamic process neglects delayed neutrons and any intrinsic reactivity feedbacks. It is mathematically described by

$$\dot{n}(t) = \frac{\rho(t)n(t)}{\ell} \quad (4-1)$$

where  $n(t)$  is the state variable, reactor power level,  $\rho(t)$  is the control reactivity, and  $\ell$  is a characteristic neutron generation time of the system. If  $\rho(t)$  were a function of  $n(t)$ , equation (4-1) would be non-linear. The technique of Klinger is used here to obtain a linear system. The pseudo-control variable is

$$u(t) = \frac{\rho(t)n(t)}{\ell} \quad (4-2)$$

such that equation (4-1) now becomes

$$\dot{n}(t) = u(t) \quad (4-3)$$

The desired quadratic error index for this example is

$$e(t) = [n(T) - n_0]^2 \phi_T + \int_t^T \left\{ \phi [n(\tau) - n_0]^2 + u(\tau)^2 \right\} d\tau \quad (4-4)$$

In equation (4-4) the control weighting factor,  $\psi$ , is chosen to be unity. Also, the control error is merely  $u(\tau)$  since, when  $n(\tau) = n_0$ , the steady state value, the control reactivity is,  $\rho(\tau) = 0$ , and thus  $u^*(\tau) = 0$ . Equation (4-4)

minimizes errors in the control,  $\rho(t)n(t)/\ell$ , and is related indirectly to minimizing errors in  $\rho(t)$  only. In the non-linear examples of this chapter the term  $u(\tau)^2$  in equation (4-4) is replaced by  $\rho(\tau)^2$ .

The necessary condition for a minimum, in the dynamic programming format, is given by equation (3-39). For this problem it is written

$$\frac{\partial E}{\partial \tau} = - \min_{u(\tau) \in U} \left\{ \phi [n(\tau) - n_0]^2 + u(\tau)^2 + u(\tau) \frac{\partial E}{\partial n} \right\} \quad (4-5)$$

The control which satisfies this equation is determined by setting the partial derivative with respect to  $u(\tau)$  equal to zero and solving for it. The result is

$$u(\tau) = - \frac{1}{2} \frac{\partial E[n(\tau), \tau]}{\partial n(\tau)} \quad (4-6)$$

and equation (4-5) becomes,

$$\frac{\partial E}{\partial \tau} + \phi [n(\tau) - n_0]^2 - \frac{1}{4} \left[ \frac{\partial E}{\partial n} \right]^2 = 0 \quad (4-7)$$

As demonstrated in chapter 3, the solution of this equation is

$$E[n, \tau] = k - 2k_1 n + k_{11} n^2 \quad (4-8)$$

Equating coefficients of the powers of  $n$  to zero after substituting this solution into equation (4-7), the following equations result:

$$\dot{k} = k_1^2 - \phi n_o^2 \quad (4-9)$$

$$\dot{k}_1 = -\phi n_o + k_1 k_{11} \quad (4-10)$$

and

$$\dot{k}_{11} = k_{11}^2 - \phi \quad (4-11)$$

As mentioned in chapter 3, only equation (4-11) need be evaluated, since  $k_1 = k_{11} n_o$ , and the boundary condition due to impulse weighting at  $\tau = T$  is

$$k_{11}(T) = \phi_T \quad (4-12)$$

The solution of equation (4-11) is

$$k_{11}(\tau) = \sqrt{\phi} \left\{ \tanh[\sqrt{\phi}(T-\tau)] + \tanh^{-1} \frac{\phi_T}{\sqrt{\phi}} \right\} \quad (4-13)$$

The resulting feedback control then is

$$u(\tau) = \sqrt{\phi} \tanh \left[ \sqrt{\phi}(T-\tau) + \tanh^{-1} \frac{\phi_T}{\sqrt{\phi}} \right] [n_o - n(\tau)] \quad (4-14)$$

Thus the optimal feedback control is a time-varying gain times the linear deviation of the power from the nominal steady state  $n_o$ . This is qualitatively as expected, for when  $n > n_o$  negative control is applied to return the power to steady state, and for  $n < n_o$  positive control is applied. Impulse weighting is seen to have the effect of increasing the control period. This effective control period is written

$$T_{\text{eff}} = T + \frac{1}{\sqrt{\phi}} \tanh^{-1} \frac{\phi_T}{\sqrt{\phi}} \quad (4-15)$$

The selection of  $\phi$  is determined by the method outlined in the previous chapter. With  $\psi = 1$ , the weighting factor is determined from the equation

$$\phi = \frac{[\delta u_{MA}]^2}{[\delta n_{MA}]^2} \quad (4-16)$$

Suppose, for example, that the maximum allowable deviation in power is the fraction  $P$  of the steady state value  $n_0$  and that the maximum available control is limited by a period constraint on the reactor. The period constraint for this model is written

$$\omega_m = (\dot{n}/n)_{\max} = (\rho(t)/\ell)_{\max} \quad (4-17)$$

and the maximum available control becomes

$$u_{\max} = \omega_m n_{\max} \quad (4-18)$$

Using the considerations that

$$\delta n_{MA} = P n_0 \quad (4-19)$$

$$\text{and } n_{\max} \text{ allowable} = (1+P)n_0 \quad (4-20)$$

the weighting factor is found to be

$$\phi = \frac{u_{\max}^2}{(P n_0)^2} = \omega_m^2 \left[ \frac{1+P}{P} \right]^2 = \omega_m'^2 \quad (4-21)$$

where  $\omega_m'$  is a weighted maximum allowable inverse period.

A plot of the feedback gain, represented by equation (4-13), as a function of  $t/T$ , for several values of  $\phi_T$ , is presented in figure 5. The maximum value of  $\phi_T$ , is  $\phi_{Tmax} = \sqrt{\phi} = \omega'_m$ . Where  $\phi_T = \omega'_m$ , the gain  $k_{11}$  is a constant for all time. For  $\phi_T < \omega'_m$  the gain  $k_{11}$  is seen to be a function of  $\omega'_m T$ , which is approximately equal to the number of relaxation times for which control effort is applied. The terminal value of the gain is seen to be a direct function of the factor  $\phi_T/\omega'_m$ .

The optimal reactor power response is given by equations (4-22) and (4-23) for the cases where

$\phi_T < \omega'_m$  and  $\phi_T = \omega'_m$  respectively

$$n(\tau) = n_o + [n(0) - n_o] \frac{\cosh[\omega'_m(T - \tau) + \tanh^{-1} \phi_T/\omega'_m]}{\cosh[\omega'_m T + \tanh^{-1} \phi_T/\omega'_m]} \quad (4-22)$$

and 
$$n(\tau) = n_o + [n(0) - n_o] e^{-\omega'_m \tau} \quad (4-23)$$

where  $n(0)$  is the initial perturbation at  $t = 0$ . Substituting these relations into the optimal control equation (4-14), the control is represented as an explicit function of time.

For  $\phi_T < \omega'_m$

$$u(\tau) = \frac{\omega'_m \sinh[\omega'_m(T - \tau) + \tanh^{-1} \phi_T/\omega'_m]}{\cosh[\omega'_m T + \tanh^{-1} \phi_T/\omega'_m]} [n(0) - n_o] \quad (4-24)$$

or for  $\phi_T = \omega'_m$

$$u(\tau) = -\omega'_m [n(0) - n_o] e^{-\omega'_m \tau} \quad (4-25)$$



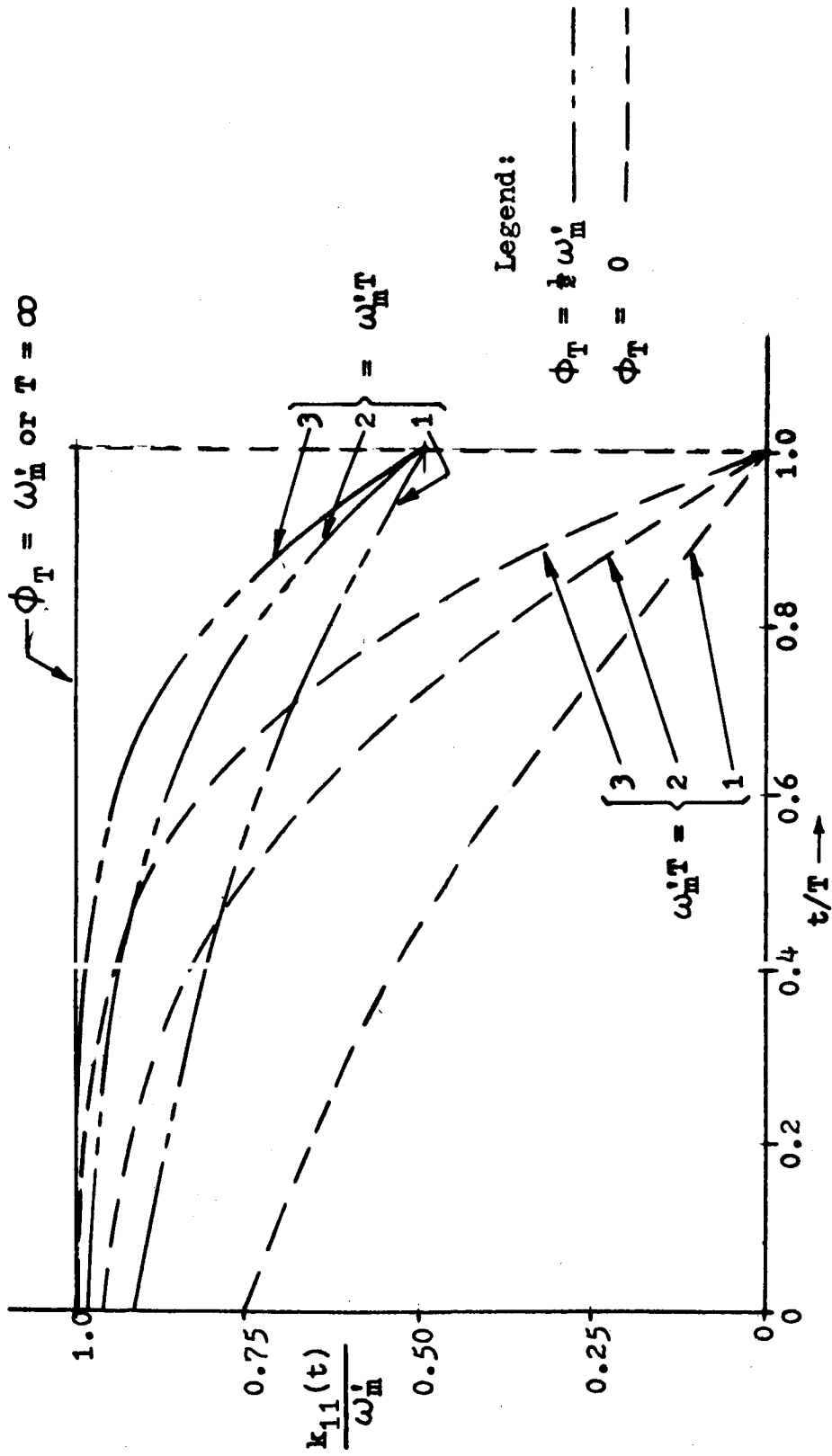


Figure 5  
Optimal Feedback Gains for a Prompt-Reactor Model

The corresponding optimal reactivities for the two cases are

$$\rho(\tau) = - \ell \omega_m' [n(0) - n_0] \sinh \left[ \omega_m' (T - \tau) + \tanh^{-1} \frac{\phi_T}{\omega_m'} \right] * \\ \left\{ n_0 \cosh \left[ \omega_m' T + \tanh^{-1} \frac{\phi_T}{\omega_m'} \right] + \right. \\ \left. [n(0) - n_0] \cosh \left[ \omega_m' (T - \tau) + \tanh^{-1} \frac{\phi_T}{\omega_m'} \right] \right\}^{-1} \quad (4-26)$$

or for  $\phi_T = \omega_m'$

$$\rho(\tau) = \frac{-\ell \omega_m'}{n_0 e^{\omega_m' \tau} + 1} \frac{1}{[n(0) - n_0]} \quad (4-27)$$

All of these equations correspond to the problem where the terminal state is not specified (the free-point terminal-boundary-value problem). In the case where the terminal power level is fixed at  $n(T) = n_0$  this constraint may be incorporated into the Hamiltonian by the use of a Lagrange multiplier. The Hamiltonian is written

$$H = \phi [n(\tau) - n_0]^2 + u(\tau)^2 + \lambda u(\tau) + u(\tau) \frac{\partial E}{\partial n} = 0 \quad (4-29)$$

The optimal control equation becomes

$$2u(\tau) + \lambda + \frac{\partial E}{\partial n} = 0 \quad (4-30)$$

The solution is assumed to have the form

$$E = k - 2k_1n + k_{11}n^2 \quad (4-31)$$

It is necessary to solve:

$$-\dot{k} = \phi n_0^2 - \frac{1}{4} \lambda^2 + \lambda k_1 - k_1^2 \quad (4-32)$$

$$-\dot{k}_1 = \phi n_0 - \frac{\lambda k_{11}}{2} - k_1 k_{11} \quad (4-33)$$

$$-\dot{k}_{11} = \phi - k_{11}^2 \quad (4-34)$$

with the boundary conditions

$$k(T) = k_1(T) = k_{11}(T) = 0 \quad (4-35)$$

Here impulse weighting is not required, due to the fact that the terminal error is constrained to be zero in the Hamiltonian. Only  $k_1$  and  $k_{11}$  are required to find the optimal control. The solutions are

$$k_{11}(\tau) = \sqrt{\phi} \tanh[\sqrt{\phi} (T-\tau)] \quad (4-36)$$

$$k_1(\tau) = \sqrt{\phi} n_0 \tanh[\sqrt{\phi} (T-\tau)] + \frac{\lambda}{2} \left\{ 1 - \operatorname{sech}[\sqrt{\phi} (T-\tau)] \right\} \quad (4-37)$$

The value of the Lagrange multiplier is found by combining equations (4-30), (4-36) and (4-37) with the boundary condition that  $n(T) = n_0$ . The result is

$$\lambda = -2 \sqrt{\phi} [n(0) - n_0] \left\{ \frac{\operatorname{sech}[\sqrt{\phi}(T)]}{\tanh[\sqrt{\phi}(T)]} \right\} \quad (4-38)$$

From equation (4-30) the value of the optimal control for a fixed-point boundary condition is then

$$u(\tau) = \sqrt{\phi} \coth[\sqrt{\phi}(T-\tau)] [n_0 - n(\tau)] \quad (4-39)$$

or in terms of the period limitation

$$u(\tau) = \omega'_m \coth[\omega'_m(T-\tau)] [n_0 - n(\tau)] \quad (4-40)$$

The power level response as a function of time is given by the relation

$$n(\tau) = n_0 + [n(0) - n_0] \frac{\sinh[\omega'_m(T-\tau)]}{\sinh[\omega'_m(T)]} \quad (4-41)$$

and the optimal control reactivity is

$$\rho(\tau) = - \frac{[n(0) - n_0] \ell \omega'_m \cosh[\omega'_m(T-\tau)]}{n_0 \sinh \omega'_m T + [n(0) - n_0] \sinh[\omega'_m(T-\tau)]} \quad (4-42)$$

Figure 6 compares the optimal power level response of equations (4-22), (4-23), and (4-41) for the fixed-point and free-point terminal-boundary conditions. Figure 7 compares the optimal reactivities corresponding to each of these responses.

If the control interval is allowed to approach infinity the solution of the feedback gains are obtained from their steady state solutions. Only the free-point

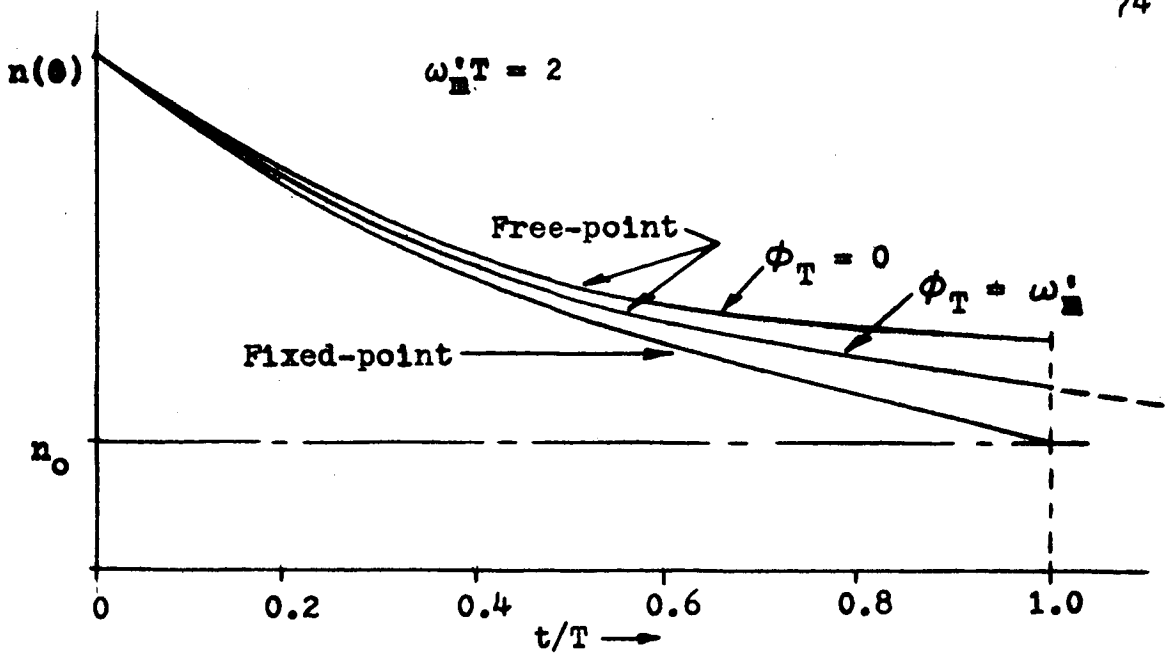


Figure 6

Comparison of Power Responses for Different Feedback Criteria

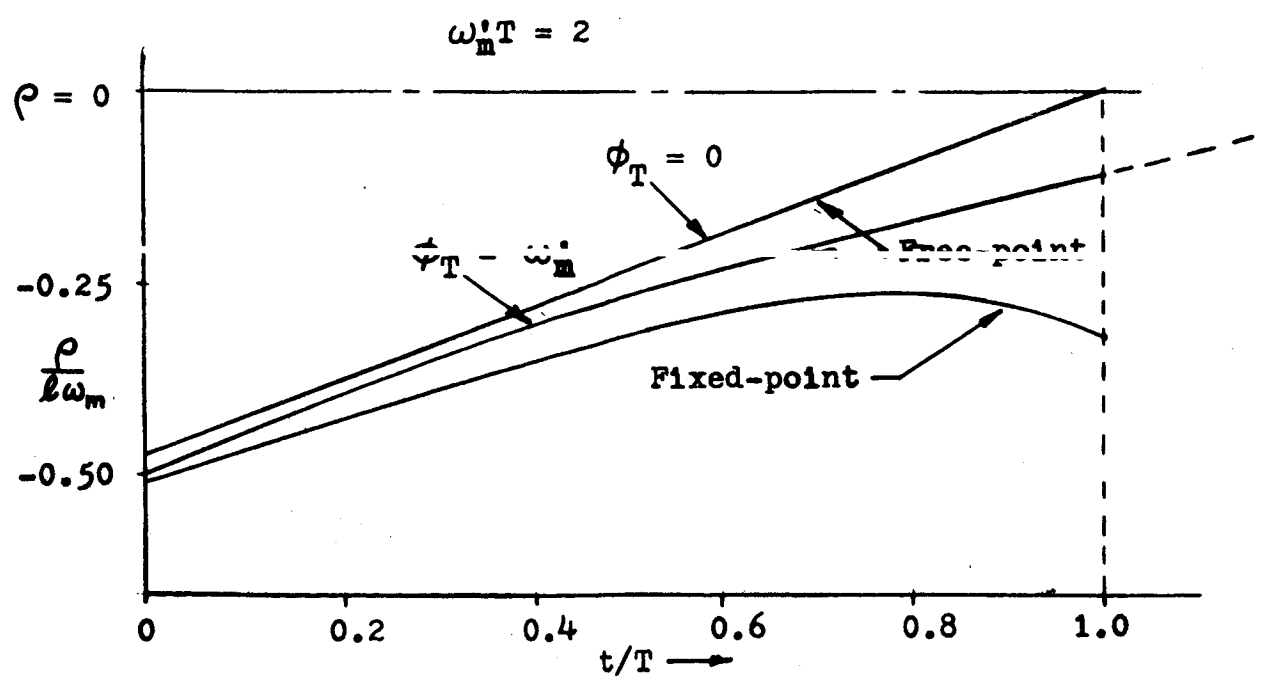


Figure 7

Compensating Reactivity for Different Feedback Criteria

condition is considered. This is seen to be identical with the free-point case where  $\phi_T = \omega_m'$  as expected. Equations (4-23) and (4-27) correspond to the optimal power level response and the feedback reactivity in this situation. The dashed extensions to the curves of figures 6 and 7 represent these conditions.

Thus far the question of saturation has not been considered. In other words, under certain conditions the maximum inverse period could be exceeded with the feedback control systems already determined. In first-order examples the treatment of saturation is not difficult for the infinite interval solution. In this case, as pointed out in chapter 2, the minimum Hamiltonian is a constant. The optimal control is the control which remains in the admissible control space and maintains a constant Hamiltonian. The result for this case is

$$u(\tau) = \begin{cases} -u_{\max} & ; \quad n(\tau) > (1+P)n_0 \\ \omega_m' [n_0 - n(\tau)] & ; \quad (1-P)n_0 \leq n(\tau) \leq (1+P)n_0 \\ u_{\max} & ; \quad n(\tau) < (1-P)n_0 \end{cases} \quad (4-43)$$

The corresponding optimal reactivity is

$$\rho(\tau) = \begin{cases} -(1+P)n_0/n(\tau) & ; \quad n(\tau) > (1+P)n_0 \\ \frac{(1+P)}{P} \left[ \frac{n_0}{n(\tau)} - 1 \right] & ; \quad (1-P)n_0 \leq n(\tau) \leq (1+P)n_0 \\ \frac{(1+P)n_0}{n(\tau)} & ; \quad n(\tau) < (1-P)n_0 \end{cases} \quad (4-44)$$

Figure 8 is a plot of the optimal reactivity for a maximum allowable perturbation in power of  $\pm 50\%$ , i.e.  $P = 0.5$ , as a function of  $n(\tau)/n_0$ .

In this rather highly oversimplified example of a nuclear reactor model a great deal of insight to the problem of optimal feedback control synthesis has been gained. It is not difficult to see that the extension to higher-order examples, or even first-order non-linear problems, could be a cumbersome task. Most of the design considerations have been demonstrated in this example.

#### Non-Linear Prompt Neutron Reactor Synthesis

The previous example will now be used to demonstrate the techniques of quasi-optimal control for non-linear dynamic processes. In this case, however, reactivity is considered as the control variable. The reactor model, here repeated, is

$$\dot{n}(t) = \frac{\rho(t)n(t)}{\ell} \quad (4-45)$$

The quadratic performance index is chosen to be

$$e(\tau) = \phi_T [n(T) - n_0]^2 + \int_t^T \left\{ \phi [n(\tau) - n_0]^2 + \rho(\tau)^2 \right\} d\tau \quad (4-46)$$

Note that reactivity errors are weighted directly here.

In this, and all first-order examples to follow, the problem will be confined to an infinite control interval where  $T = \infty$ . This is not necessary for non-linear control,

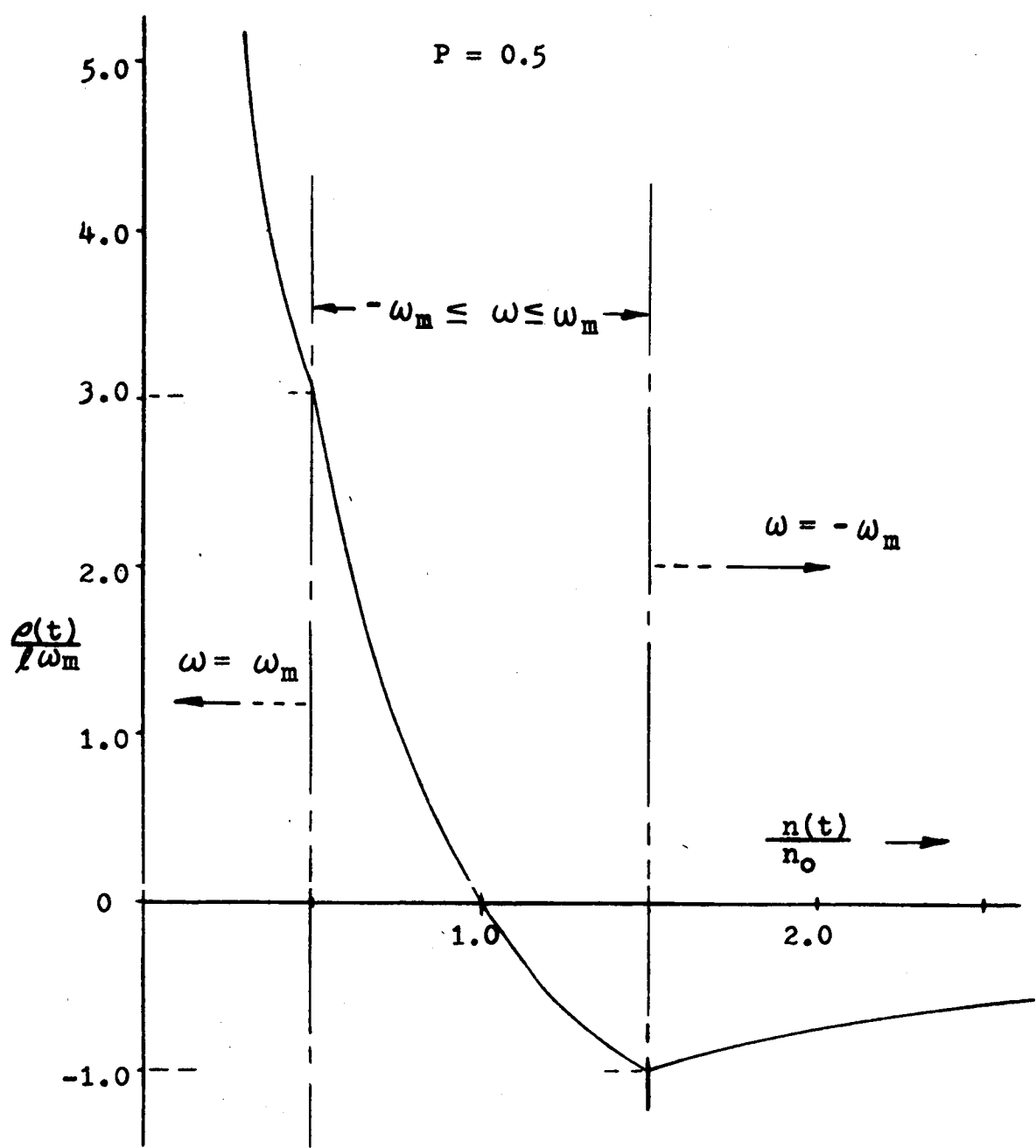


Figure 8

Optimal Control Reactivity with Minimum Period Constraint



but when a reactor is operating in the steady state, it is usually desirable to do so for extended lengths of time. In such cases the control interval is much much larger than the characteristic time constants of the system.

For this example then, the exact non-linear optimal control is solvable using the results presented in equations (2-32) through (2-36). The error criterion becomes

$$e(\tau) = \int_t^{\infty} H[n, \rho] d\tau = \int_t^{\infty} \left\{ \phi [n(\tau) - n_0]^2 + \rho(\tau)^2 \right\} d\tau \quad (4-47)$$

From equation (2-36) the optimal value of the time derivative of the reactor power is

$$\dot{n}^*(\tau) = \frac{H - C}{\partial H / \partial \dot{n}} = \frac{\rho^*(\tau) n^*(\tau)}{\ell} \quad (4-48)$$

When the control interval is infinite, the terminal value of the power level is equal to the nominal value,  $n_0$ , and  $\rho(\tau) = 0$ . Thus the constant of integration  $C$  vanishes as seen from equation (4-47). The derivative at any time is then

$$\dot{n}(\tau) = \frac{H}{\partial H / \partial \dot{n}} \quad (4-49)$$

where

$$H = \phi [n - n_0]^2 + \frac{\dot{n}^2 \ell^2}{n^2} \quad (4-50)$$

using the definition of  $\rho(\tau)$  in equation (4-47).

Taking the partial derivative and solving for  $\dot{n}(\tau)$

$$\dot{n}(\tau) = \pm \left| \frac{n(\tau)[n(\tau)-n_0]\sqrt{\phi}}{\ell} \right| \quad (4-51)$$

The exact optimal reactivity for the infinite interval control then is

$$\rho(\tau) = \sqrt{\phi} [n_0 - n(\tau)] \quad (4-52)$$

In this special case, the optimal feedback reactivity is a linear function of the deviation of the power from the nominal value.

This is a very simple control to synthesize, but in order to demonstrate quasi-optimal techniques, the methods of Pontryagin and Bellman are applied and equation (4-52) is used as a basis for comparison. First, Pontryagin's maximum principle is considered.

The perturbed state and control variables for this case are

$$\delta\rho(\tau) = \rho(\tau) \text{ and } \delta n(\tau) = n(\tau) - n_0 \quad (4-53)$$

The linearized differential equation that describes small perturbations about the nominal values of  $n = n_0$  and  $\rho = 0$ , is

$$\delta\dot{n} = \frac{n_0 \delta\rho}{\ell} \quad (4-54)$$

and the constrained Hamiltonian is

$$H_c = \phi\delta n^2 + \delta\rho^2 + \frac{\lambda n_0 \delta\rho}{\ell} \quad (4-55)$$

The necessary conditions for a minimum are

$$\frac{\partial H_c}{\partial \delta \rho} = 0 = 2 \delta \rho + \frac{\lambda n_0}{\ell} \quad (4-56)$$

and  $\frac{\partial H_c}{\partial \delta n} = -\dot{\lambda} = 2 \phi \delta n \quad (4-57)$

Now  $\lambda(\tau)$  is assumed to be of the form

$$\lambda(\tau) = k(\tau)n(\tau) \quad (4-58)$$

and  $k(\tau)$  is found from the relation

$$-\dot{k}(\tau) = -2\phi + \frac{k(\tau)^2 n_0^2}{2\ell^2} \quad (4-59)$$

The steady state solution is required for the infinite interval problem and thus

$$k(\tau) = \pm \frac{2\ell\sqrt{\phi}}{n_0} \quad (4-60)$$

Combining equations (4-56), (4-58) and (4-60) the quasi-optimal reactivity of the maximum principle is

$$\delta \rho(\tau) = \rho(\tau) = \sqrt{\phi} [n_0 - n(\tau)] \quad (4-61)$$

This is seen to be identical with the exact optimal control of equation (4-52). This is as expected, since the maximum principle yields a quasi-optimal control which is truncated at the first-power of  $\delta n(\tau)$  and equation (4-52) is linear in  $\delta n(\tau)$ .

Since either the exact or the maximum principle give identical synthesis controllers, it would be unnecessary to consider a quasi-optimal control based on dynamic programming. However, in this example additional results are found using dynamic programming. Also, in order to demonstrate the technique, this controller is evaluated for any arbitrary  $p^{\text{th}}$ -degree expansion of the minimum error function. Equations (3-95) through (3-103) gives the required format of this technique.

The  $p^{\text{th}}$ -degree expansion of the minimum error function is

$$E_p = k_0 - 2k_1n + k_2n^2 + \dots + \frac{2}{p} k_p n^p \quad (4-62)$$

The  $p^{\text{th}}$ -degree Hamiltonian is

$$H_p^* = \phi [n(\tau) - n_0]^2 + \rho(\tau)^2 + \frac{\rho(\tau)n(\tau)}{\ell} \left[ \frac{\partial E_p}{\partial n} \right] \quad (4-63)$$

The control which minimizes (4-63) is

$$\rho_p(\tau) = - \frac{1}{2} \frac{n(\tau)}{\ell} \left[ \frac{\partial E_p}{\partial n} \right] \quad (4-64)$$

Thus minimum Hamiltonian is

$$H_p^{**} = \phi [n(\tau) - n_0]^2 - \frac{n(\tau)^2}{4\ell^2} \left[ \frac{\partial E_p}{\partial n} \right]^2 \quad (4-65)$$

$$= H_0^* + H_1^* n + H_2^* n^2 + \dots + H_p^* n^p \quad (4-66)$$

For the infinite interval problem the  $k_j$  functions are all equal to zero and thus  $H_j^*$  are all zero also. The  $H_j^*$  functions are defined by equation (3-99).

For a first-degree expansion of  $p = 1$

$$E_1 = k_0 - 2k_1 n \quad (4-67)$$

$$\rho_1 = \frac{k_1 n}{\ell} \quad (4-68)$$

and 
$$H_1^{i*} = \phi(n-n_0)^2 - \frac{n^2 k_1^2}{\ell^2} \quad (4-69)$$

It is not necessary to find the  $H_0^*$  function, because the feedback control does not depend on it. The value of  $k_1$  is found from

$$H_1^* = 0 = \left[ \frac{\partial H_1^{i*}}{\partial n} \right]_{n=n_0} = \left[ 2\phi(n-n_0) - \frac{2nk_1^2}{\ell^2} \right]_{n=n_0} \quad (4-70)$$

Thus  $k_1 = 0$  and  $\rho_1 = 0$  for all time.

For the second-degree expansion

$$E_2 = k_0 - 2k_1 n + k_2 n^2 \quad (4-71)$$

$$\rho_2 = \frac{k_1 n}{\ell} - \frac{k_2 n^2}{\ell} \quad (4-72)$$

and 
$$H_2^{i*} = \phi(n-n_0)^2 - \frac{n^2}{\ell^2} [k_1^2 - 2k_1 k_2 n + k_2^2 n^2] \quad (4-73)$$

The equations necessary to evaluate the gains  $k_1$  and  $k_2$  are

$$-\ell^2 \phi n_0 + 3k_1 k_2 n_0^2 + 4k_2^2 n_0^3 = 0 \quad (4-74)$$

$$\text{and } -\ell^2 \phi + k_1^2 - 6k_1 k_2 n_0 + 6k_2^2 n_0^2 = 0 \quad (4-75)$$

From the above equations it is seen that

$$k_1 = k_2 n_0 \quad (4-76)$$

and the solution is

$$k_1 = \ell \sqrt{\phi} \quad \text{and} \quad k_2 = \ell \sqrt{\phi} / n_0 \quad (4-77)$$

The quasi-optimal control then is

$$\rho_2(\tau) = \sqrt{\phi} \left[ \frac{n - \frac{n^2}{n_0}}{n_0} \right] \quad (4-78)$$

It can readily be shown that the general expression for the reactivity for the  $p^{\text{th}}$ -degree expansion is

$$\rho_p(\tau) = n_0 \sqrt{\phi} \left[ 1 - \frac{n(\tau)}{n_0} - \left( 1 - \frac{n(\tau)}{n_0} \right)^p \right] \quad (4-79)$$

The weighting factor  $\phi$  is determined to be

$$\phi = \left[ \frac{\rho_{\max}}{P n_0} \right]^2 \quad (4-80)$$

Here a maximum reactivity is assumed to be the constraint rather than a period constraint, because the error criterion directly weights reactivity. Equation (4-79) is now written

$$\rho_p(\tau) = \frac{\rho_{\max}}{P} \left[ 1 - \frac{n(\tau)}{n_0} - \left( 1 - \frac{n(\tau)}{n_0} \right)^p \right] \quad (4-81)$$

A comparison of the quasi-optimal reactivity of equation (4-81) for several degree expansions with the exact optimal reactivity (also the maximum principle) is indicated in figure 9. These curves indicate the regions of validity of the quasi-optimum control. For odd-degree expansions, equation (4-81) is not valid for perturbations greater than twice the nominal value, because positive reactivity results. For even-degree expansions the approximate control deviates far from the optimal for deviations greater than  $2n_0$ , and become very large negative. The range of validity is restricted to

$$0 < n(\tau) < 2n_0 \quad (4-82)$$

The equal signs are not included, because for these values unstable transitions occur, at least for odd-degree expansions. In the range of validity, convergence to the optimal is achieved for each successive expansion.

Whereas this example illustrates that an unstable control is achieved from the dynamic programming approach, the other non-linear examples in this chapter show that the quasi-optimal control of dynamic programming is not only stable, but superior with respect to accuracy. Not only is the dynamic programming example undesirable with regard to accuracy, but the complexity of the feedback controller is greater than for the maximum principle. Figure 10 illustrates the controllers for equation (4-61) and (4-78).

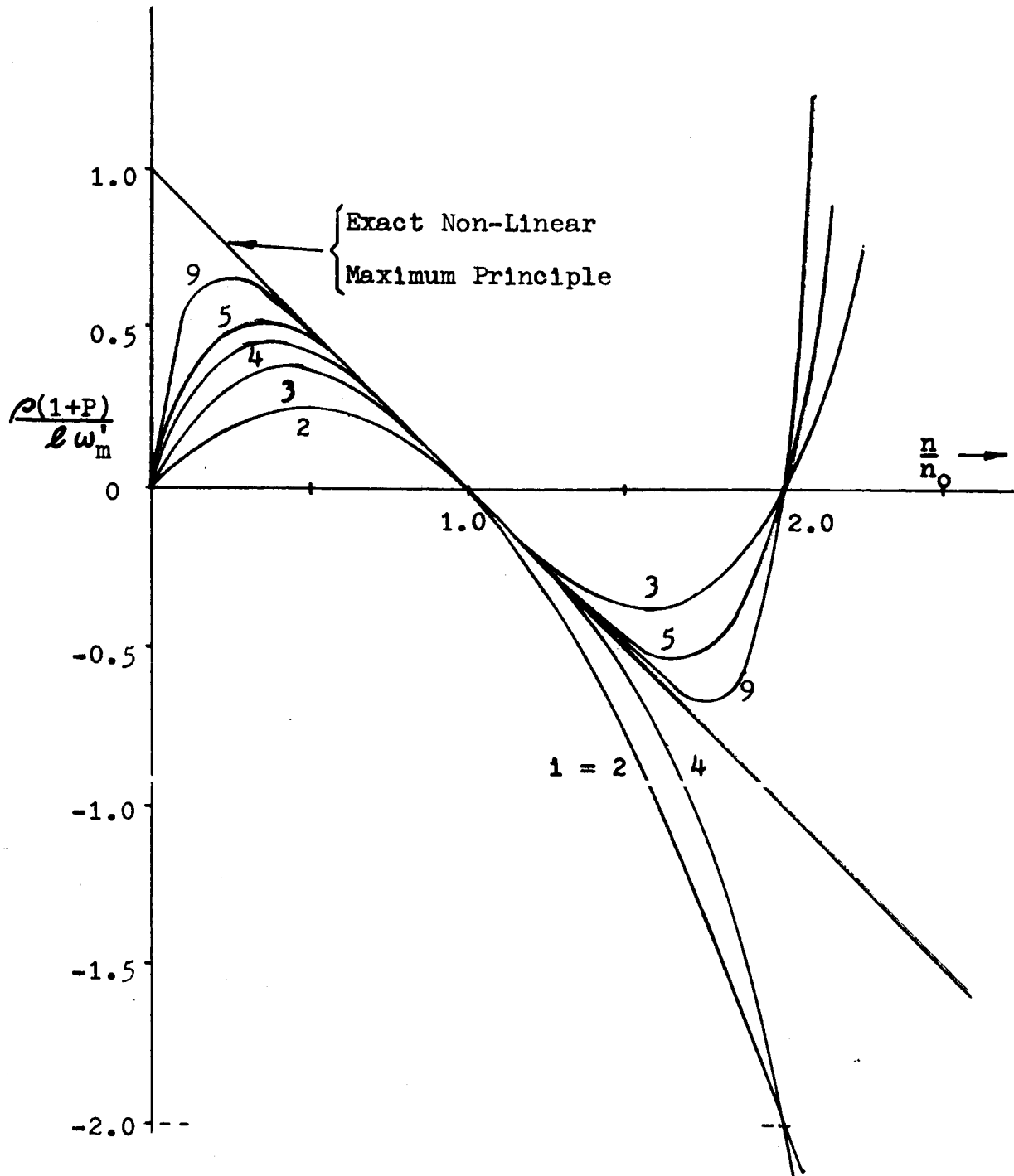
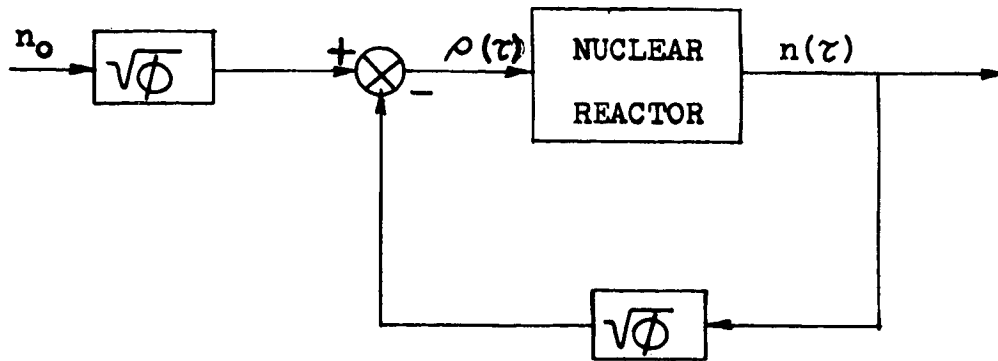


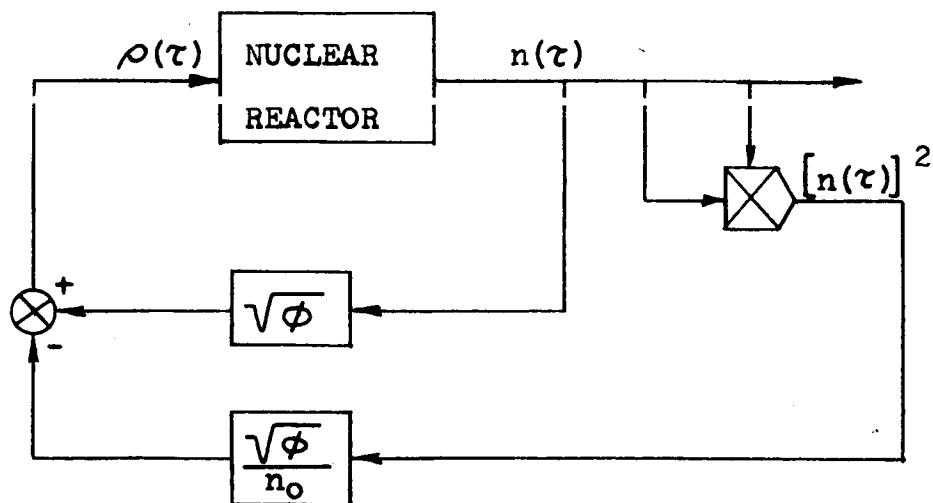
Figure 9

Quasi-Optimal Control vs Exact Non-Linear Optimal





(a) Optimal Feedback Controller



(b) Quasi-Optimal Feedback Controller

Figure 10

Feedback Controllers for Non-Linear Reactor Model

### Linear Delayed Neutron Reactor Synthesis

The reactor dynamic process chosen here includes an intrinsic negative feedback mechanism representative of the delayed neutron effect. Steady state operation is assumed, and the pseudo-control variable,  $u = \rho n/\ell$ , is used. This model is represented

$$\dot{n}(\tau) = u(\tau) + \frac{\beta}{\ell} [n_0 - n(\tau)] \quad (4-83)$$

The error criterion chosen is

$$e(t) = \int_t^{\infty} \left\{ \phi [n(\tau) - n_0]^2 + u(\tau)^2 \right\} d\tau \quad (4-84)$$

The necessary condition for a minimum control is written

$$\frac{\partial E}{\partial \tau} + \min_{u(\tau)} \left\{ \phi [n(\tau) - n_0]^2 + u(\tau)^2 + \left[ u(\tau) + \frac{\beta}{\ell} [n_0 - n(\tau)] \right] \frac{\partial E}{\partial n} \right\} = 0 \quad (4-85)$$

The optimal control is determined from equation (4-85) to be

$$u(\tau) = - \frac{1}{2} \frac{\partial E}{\partial n} \quad (4-86)$$

and the solution, as before, is

$$E(n, \tau) = k_2(\tau) [n(\tau) - n_0]^2 \quad (4-87)$$

The gain  $k_2(\tau)$  satisfies

$$\dot{k}_2(\tau) = k_2^2(\tau) + \frac{2\beta}{\ell} k_2(\tau) - \phi \quad (4-88)$$

The steady state solution is sufficient and thus  $\dot{k}_2(\tau) = 0$ .

The result is

$$k_2 = -\frac{\beta}{l} + \sqrt{\left(\frac{\beta}{l}\right)^2 + \phi} \quad (4-89)$$

and the optimal feedback control, from equations (4-86), (4-87) and (4-89), is

$$u(\tau) = -\left[\frac{\beta}{l} - \sqrt{\left(\frac{\beta}{l}\right)^2 + \phi}\right][n_0 - n(\tau)] \quad (4-90)$$

Here again,  $\phi$  is determined from the period constraint and the maximum allowable deviation in power,  $n_{\max} = (1+P)n_0$ .

$$\phi = \frac{u_{\max}^2}{(Pn_0)^2} \quad (4-91)$$

For this case  $u_{\max}$  is written

$$\begin{aligned} u_{\max} &= \left(\frac{\dot{n}}{n}\right)_{\max} n_{\max} - \frac{\beta}{l}(n_0 - n_{\max}) \\ &= n_0 \left[ (1+P)\omega_m + \frac{P\beta}{l} \right] \end{aligned} \quad (4-92)$$

The amount of control reactivity to maintain  $\dot{n} = 0$  for an impulse in power of  $(1+P)n_0$  is roughly  $\rho_{\max} = P\beta$ . Thus

$$\omega_m = \frac{\rho_{\max}}{l} = \frac{P\beta}{l} \quad (4-93)$$

Substituting equations (4-92) and (4-93) into equation (4-91) the weighting factor is written

$$\phi = (4 + 2P + P^2) \frac{\beta^2}{l^2} \quad (4-94)$$

The optimal feedback control for equation (4-90) is then written

$$u(\tau) = \frac{\rho_{\max}}{\ell P} \left[ \sqrt{5+2P+P^2} - 1 \right] [n_0 - n(\tau)] \quad (4-95)$$

A new variable, corresponding to a weighted inverse period, is introduced by

$$u(\tau) = \omega_m'' [n_0 - n(\tau)] \quad (4-96)$$

The optimal feedback reactivity is

$$\rho(\tau) = \omega_m'' \left[ \frac{n_0}{n(\tau)} - 1 \right] \quad (4-97)$$

Equation (4-97) is seen to be very similar to the reactivity for a prompt neutron model of equation (4-44) if  $\ell$  is equivalent in both models. With delayed neutrons the feedback reactivity differs by a factor here defined as  $\Omega(P)$ . Equation (4-97) is rewritten

$$\rho(\tau) = \omega_m' \Omega(P) \left[ \frac{n_0}{n(\tau)} - 1 \right] \quad (4-98)$$

where

$$\Omega(P) = \frac{\omega_m''}{\omega_m'} = \frac{(P + 1)}{\sqrt{5+2P+P^2} - 1} \quad (4-99)$$

Figure 11 is a plot of  $\Omega(P)$  versus  $P$  and shows that it never deviates by more than 20 per cent from unity. This indicates that the linear optimal control for prompt neutrons only, very nearly represents the optimal control when delayed neutrons are considered regardless of the value of  $\beta$ . The

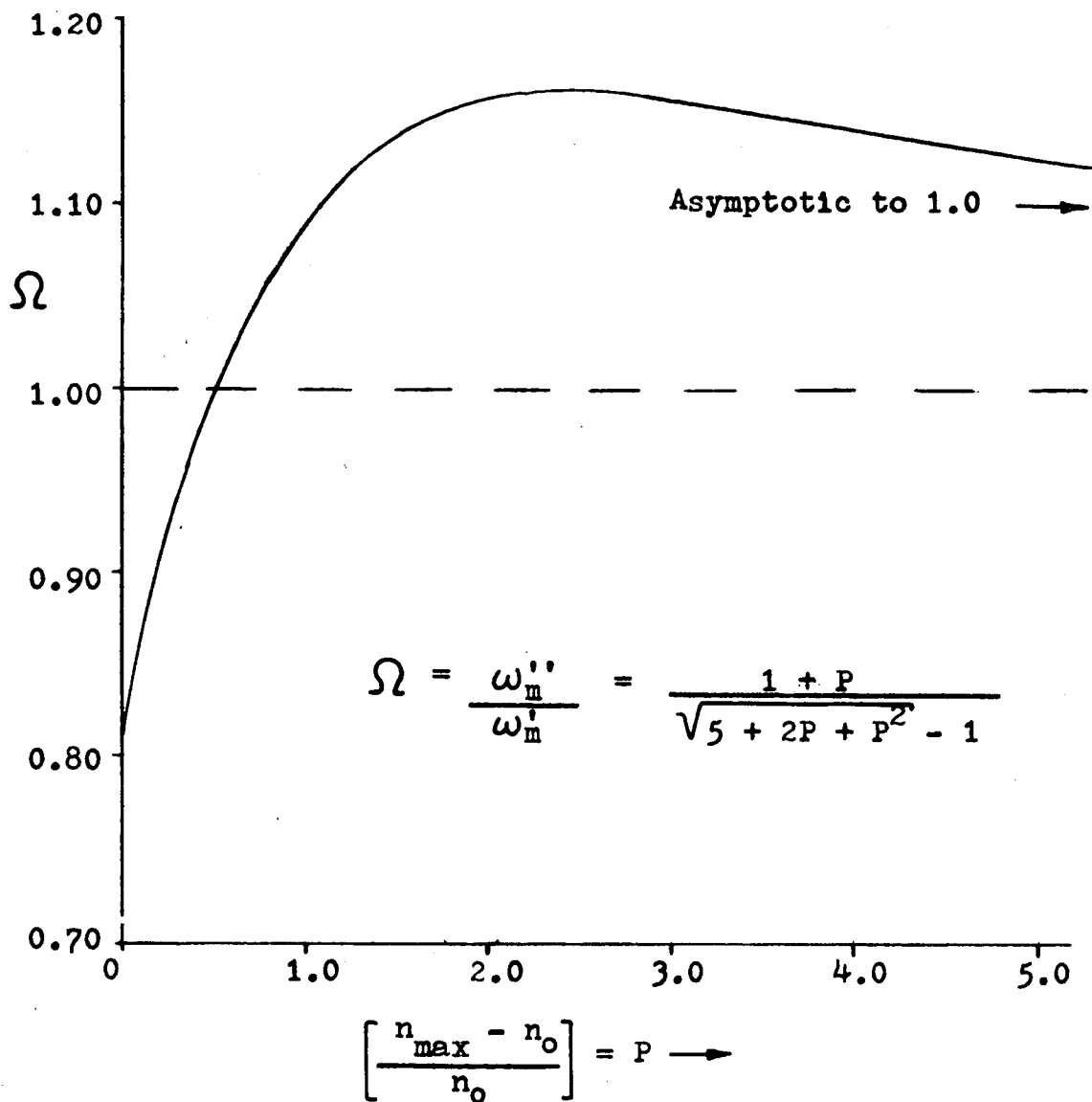


Figure 11

Ratio of Inverse Periods vs Allowable Power Deviation

only variables that need be specified are the minimum reactor period and the maximum allowable deviation in power.

An important result here is that the measurement of reactor power is the only state required, eliminating the need to calculate precursor densities. Such a feedback controller appears ideal for steady state operation of power reactors where internal noise and small external fluctuating loads occur. Boiling water, pressurized water, or even SNAP reactors are examples of this case.

#### Non-Linear Delayed Neutron Reactor Synthesis

Here, again, the model represents the effect of delayed neutrons, but the error index weights errors in reactivity separately, rather than the control  $\rho n/l$ . This is the non-linear synthesis problem requiring the quasi-optimal techniques already demonstrated. The dynamic process is written

$$\dot{n}(\tau) = \frac{\rho(\tau)n(\tau)}{\bar{l}} + \frac{\beta}{\bar{l}} [n_0 - n(\tau)] \quad (4-100)$$

In this section it is demonstrated that dynamic programming quasi-optimal control of 2<sup>nd</sup>-degree is superior to the maximum principle. It is also demonstrated that a 2<sup>nd</sup>-degree Taylor expansion of the exact optimal control equation (the maximum principle is a 1<sup>st</sup>-degree expansion) yields a region of unstable control, whereas no limitations result from dynamic programming.

The error criterion for this study is

$$e(t) = \int_t^{\infty} \left\{ \phi [n(\tau) - n_0]^2 + \rho(\tau)^2 \right\} d\tau \quad (4-101)$$

In the infinite interval problem the exact optimal control, using the results presented in chapter 2, is given by

$$\rho(\tau) = \beta \left[ 1 - \frac{n_0}{n(\tau)} \right] \left[ 1 - \sqrt{\frac{\phi n(\tau)^2 + 1}{\beta^2}} \right] \quad (4-102)$$

The weighting factor  $\phi$  is determined from the reactivity constraint. If  $P$  is the maximum allowable fractional deviation in power, then  $P\beta$  is roughly the maximum compensating reactivity required to maintain  $\dot{n} = 0$ . Thus

$$\phi = \frac{(P\beta)^2}{(Pn_0)^2} = \left[ \frac{\beta}{n_0} \right]^2 \quad (4-103)$$

and the optimal reactivity is

$$\rho(\tau) = \beta \left[ 1 - \frac{n_0}{n(\tau)} \right] \left[ 1 - \sqrt{1 + \frac{n(\tau)^2}{n_0^2}} \right] \quad (4-104)$$

The maximum principle quasi-optimal control, and a 2<sup>nd</sup>-degree Taylor expansion of the optimal control equation, are directly obtainable from equation (4-104). The maximum principle quasi-optimal control equation is

$$\rho_1(\tau) = \beta [\sqrt{2}-1] \left[ 1 - \frac{n(\tau)}{n_0} \right] \quad (4-105)$$

and the 2<sup>nd</sup>-degree expansion gives

$$\rho_2(\tau) = \beta \left[ (2\sqrt{2}-3) + (5-3\sqrt{2}) \frac{n(\tau)}{n_0} + (\sqrt{2}-2) \frac{n(\tau)^2}{n_0^2} \right] \quad (4-106)$$

The dynamic programming approach of approximating the minimum error function has been carried out for three separate expansions, namely;  $P = 1, 2,$  and  $3$ . The results of these expansions for the infinite interval problem are presented here without the associated mathematical details. The three resulting quasi-optimal reactivities are

$$\rho_1 = 0 \quad (4-107)$$

$$\rho_2(\tau) = \beta \left[ \sqrt{2} - 1 \right] \left[ \frac{n(\tau)}{n_0} - \frac{n(\tau)^2}{n_0^2} \right] \quad (4-108)$$

and

$$\rho_3(\tau) = \beta \left[ \left( \frac{2-\sqrt{2}}{2} \right) \frac{n(\tau)}{n_0} + (3-2\sqrt{2}) \frac{n(\tau)^2}{n_0^2} + \left( \frac{4-3\sqrt{2}}{2} \right) \frac{n(\tau)^3}{n_0^3} \right] \quad (4-109)$$

Figure 12 is a plot of equations (4-104), (4-105), (4-106), and (4-108). Here, the comparisons are indicated between quasi-optimal control and the exact non-linear control. Several features are demonstrated in this figure. In the region where  $n/n_0$  is less than unity, the maximum principle control deviates significantly from the exact optimal, but the 2<sup>nd</sup>-degree dynamic programming control is very nearly identical. Furthermore, a control based on a 2<sup>nd</sup>-degree expansion of the exact optimal renders an unstable system for  $n/n_0$  less than  $(1 - 1/\sqrt{2})$ . The higher-degree dynamic programming controls converge rapidly to the optimal.



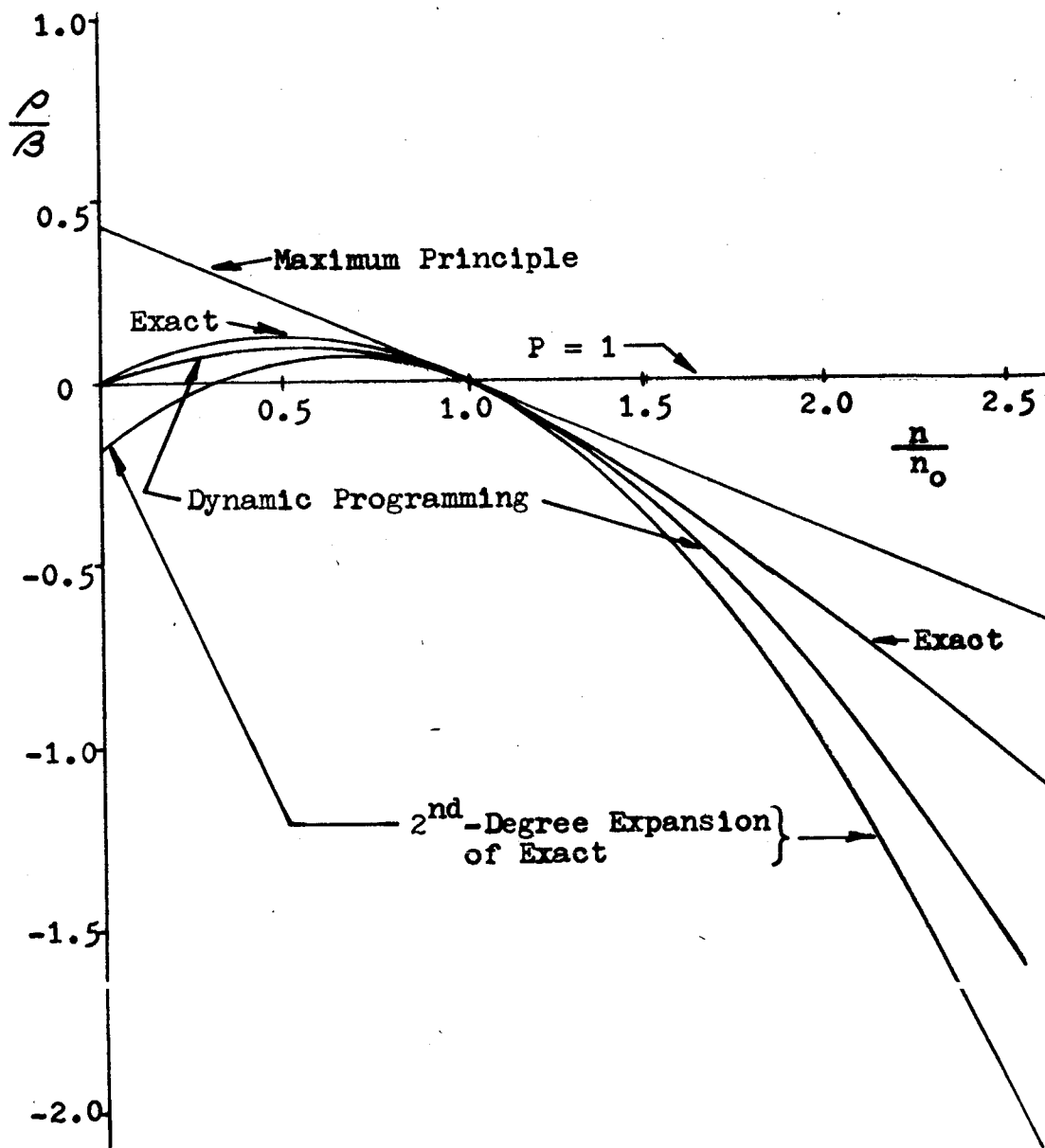


Figure 12

Comparison of Control Laws for a Non-Linear  
Reactor with Delayed Neutrons

Therefore, in this example dynamic programming yields a more nearly accurate quasi-optimal control. The maximum principle does not produce an unstable control system, however, and is less complex to synthesize. In this case also, the exact optimal control given by equation (4-104) would be very complex to synthesize in contrast to the exact non-linear control law for the prompt neutron model.

Thus, in selecting the appropriate quasi-optimal controller, the choice remains between control system complexity and desired system performance. From the two non-linear examples presented, it is evident that each control problem is unique and generalizations are not easily made. It can be said, however, that no state-determined quasi-optimal controller is less complex than one obtained from the maximum principle.

#### Another Non-Linear Reactor Model

Finally, a first-order dynamic model is considered which is non-linear and includes an additional intrinsic feedback reactivity. The purpose here is to demonstrate that the optimal (or quasi-optimal) compensating feedback reactivity is capable of maintaining a steady state operating reactor even when the intrinsic reactivity may be positive.

The non-linear model chosen incorporates a power coefficient of reactivity into the model of the last section. The model is

$$\dot{n} = \frac{\rho n}{\ell} + \frac{\alpha(n-n_0)n}{\ell} + \frac{\beta(n_0-n)}{\ell} \quad (4-110)$$

where  $\alpha(n-n_0)$  is the reactivity due to a perturbation in power. The control variable is  $\rho$ . In order to compare the optimal control of this model with that of the previous section the same quadratic error criterion is chosen, and here repeated

$$E = \int_t^{\infty} \left\{ \phi [n(\tau) - n_0]^2 + \rho^2(\tau) \right\} d\tau \quad (4-111)$$

The exact non-linear optimal control which satisfies the above error criterion is determined to be

$$\rho = \beta \left[ 1 - \frac{n_0}{n} \right] \left\{ 1 - \left( \frac{\alpha n_0}{\beta} \right) \frac{n}{n_0} - \left[ 1 - 2 \left( \frac{\alpha n_0}{\beta} \right) \frac{n}{n_0} + \left[ 1 + \left( \frac{\alpha n_0}{\beta} \right)^2 \right] \frac{n^2}{n_0^2} \right]^{\frac{1}{2}} \right\} \quad (4-112)$$

Equation (4-112) reduces to the optimal control given by equation (4-104) of the previous section when  $\frac{\alpha n_0}{\beta} = 0$ , as expected. Figure 15 is a plot of the ratio  $\rho/\beta$  from equation (4-112) for several values of  $\alpha n_0/\beta$ , both positive and negative, as a function of  $n/n_0$ . As expected, for positive feedback ( $\alpha n_0/\beta$  greater than zero), relatively more compensating control effort is required than for negative feedback.

For negative feedback the system is more stable than the model where no feedback occurs and is of little importance here. For positive feedback, however, the problem is more interesting. The quasi-optimal controls for dynamic

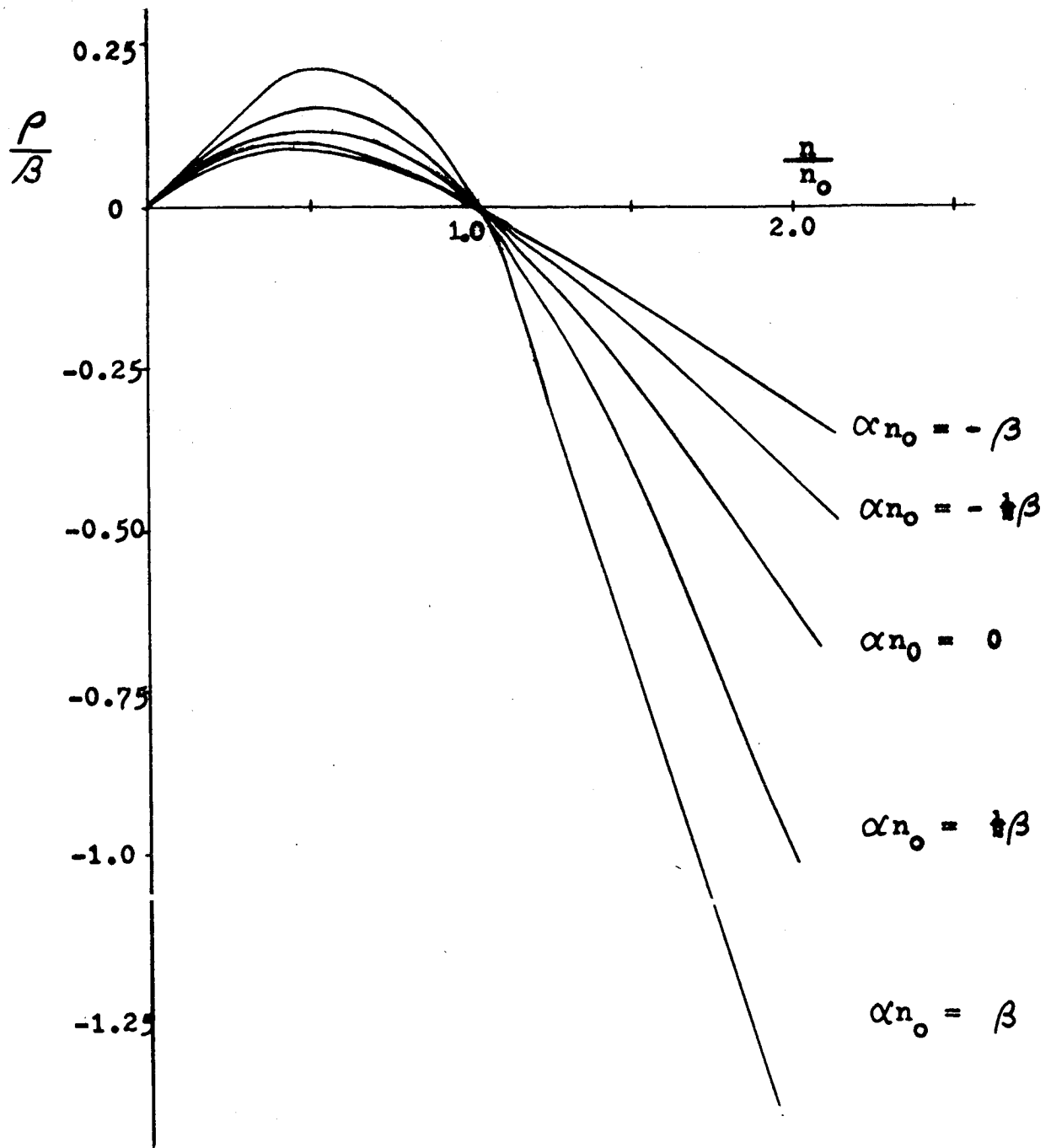


Figure 13

Exact Control for Different Power Coefficients  
of Reactivity

programming using a second-order expansion, for the maximum principle, and for a second-order expansion of equation (4-112) about  $n = n_0$  are given in equations (4-113) through (4-115) respectively. A comparison of these equations with the exact optimal for  $\alpha n_0 = \beta$  is illustrated in figure 14.

$$\rho_{DP} = \beta \left[ s - 1 + \sqrt{s^2 - 2s + 2} \right] \left[ \frac{n}{n_0} - \frac{n^2}{n_0^2} \right] \quad (4-113)$$

$$\rho_{MP} = \beta \left[ s - 1 + \sqrt{s^2 - 2s + 2} \right] \left[ 1 - \frac{n}{n_0} \right] \quad (4-114)$$

and

$$\rho_2 = \beta \left[ s - 1 + \sqrt{s^2 - 2s + 2} \right] \left[ 1 - \frac{n}{n_0} \right] \left[ 1 - \frac{2}{\sqrt{s^2 - 2s + 2}} \left( 1 - \frac{n}{n_0} \right) \right] \quad (4-115)$$

where

$$s = \frac{\alpha n_0}{\beta} \quad (4-116)$$

It is seen that the second-degree dynamic programming equation agrees well with the exact solution while equations (4-114) and (4-115) deviate quite far. The second-degree expansion of the exact optimal, given by equation (4-115), is seen to yield an unstable system for  $n$  less than  $n_0/2$ .

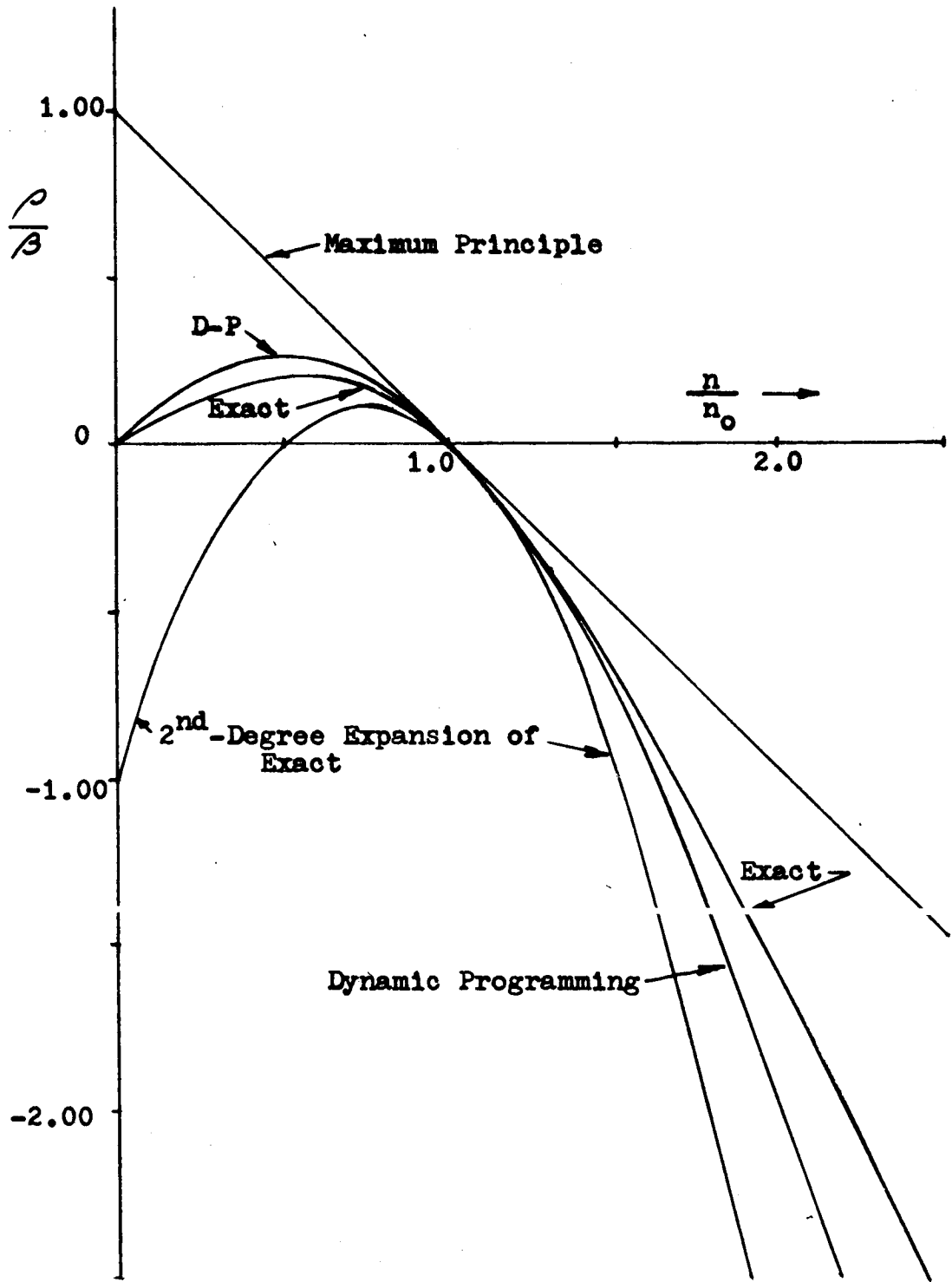


Figure 14

Comparison of Quasi-Optimal Controls  
with Exact for  $\frac{\alpha n_0}{\beta} = 1.0$

## Chapter 5

### HIGHER-ORDER NUCLEAR REACTOR SYNTHESIS

#### Introduction

The application of both linear optimal, and non-linear quasi-optimal, feedback control synthesis for three higher than first-order nuclear reactor dynamic processes is studied in this chapter. Each of these problems, when extensively evaluated, would comprise a large study. However, each problem is concerned with a different aspect of reactor control, and many facets of the control problem are demonstrated in the combined studies.

A linear reactor model, described by prompt-neutron multiplication together with delayed neutron feedback (with no other intrinsic reactivity feedback), is first considered. Optimal feedback gains are determined for the infinite-interval control problem. A one group model is used to show approximately how sensitive the equations required for a six group model are to changes in reactor lifetime.

Next, a synthesis of a nuclear reactor power transfer from 10 kilowatts to 50 kilowatts for a non-linear reactor process with intrinsic feedback reactivity proportional to power, is evaluated. A slightly different control variable, namely rate of control reactivity, is chosen. This is the first example where a finite control interval is used. The

nominal trajectories for both state and control variables are chosen to minimize the control energy during the transfer.

Finally, start-up of a nuclear rocket engine is synthesized. Several different aspects of a control problem are introduced in this example. A binary control system is needed for rocket engine synthesis. In addition to a reactivity control mechanism, the flow rate of liquid hydrogen propellant is also a control variable. The nominal control trajectories have discontinuities at three different switching times. The resulting quasi-optimal time-varying feedback gains are also discontinuous. Analog and digital computer synthesis is demonstrated for this problem.

These studies are by no means complete. However, they do effectively indicate some of the considerations required for higher-order synthesis. These are fairly realistic problems, although they are considerably simplified for this work. For example, in most situations measurement errors and large fluctuating loads may be incurred. Stochastic optimal control theory would be needed for an adequate evaluation (4). Generally the point-reactor-kinetics model is too simple to represent the overall reactor performance during dynamic operation. Kliger (6) has studied the optimal space-time-dependent reactor synthesis problem and the optimization methods presented in this work apply straightforwardly to this case. In some cases the transient response of the system may be more important, from a performance



standpoint, than the accumulated errors. Frequency-domain control synthesis can easily be incorporated in such instances.

The ultimate usefulness of optimal control theory will be decided on the basis of how easily the synthesis problem can be evaluated. In higher-order non-linear problems many special techniques must be utilized in the construction of the controller. For example, special digital or analog computer programs are required for the solution of the gain equations. A great many approximation techniques may be required. However, most of these limitations are surmountable and generally a satisfactory quasi-optimal control system results. Fortunately, a large amount of flexibility is possible with the techniques of time-domain synthesis as evidenced in this work.

#### Delayed Neutron Reactor Synthesis: Six and One Group Models

In order to introduce the application of optimal control theory to the higher order problems, a linear reactor example is chosen first. The reactor process in this section is described by a seventh-order dynamic process including state variables of reactor power level, and six delayed neutron groups. The pseudo-control  $\rho n/l$  is used. The mathematical model is

$$\dot{n} = u - \frac{\beta n}{l} + \sum_{i=1}^6 \lambda_i c_i \quad (5-1)$$

$$\dot{c}_i = \frac{\beta_i n}{l} - \lambda_i c_i \quad i = 1, 2, \dots, 6 \quad (5-2)$$

In this case a feedback controller is desired that minimizes errors in both power level and precursor densities from their nominal values. The infinite control interval error criterion may be written

$$e(t) = \int_t^{\infty} \left\{ \phi_n [n(\tau) - n^*(\tau)]^2 + \sum_{i=1}^6 \phi_{c_i} [c_i(\tau) - c_i^*(\tau)]^2 + [u(\tau) - u^*(\tau)]^2 \right\} d\tau \quad (5-3)$$

The solution to the linear optimal control problem is given by the Riccati matrix equation in chapter 3. The necessary system and input matrices for the dynamic process are

$$A = \begin{bmatrix} -\beta/l & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 & \lambda_6 \\ \beta_1/l & -\lambda_1 & & & & & \\ \beta_2/l & & -\lambda_2 & & & & \\ \beta_3/l & & & -\lambda_3 & & & \\ \beta_4/l & & & & -\lambda_4 & & \\ \beta_5/l & & & & & -\lambda_5 & \\ \beta_6/l & & & & & & -\lambda_6 \end{bmatrix} \quad B = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ v \\ 0 \\ 0 \end{bmatrix} \quad (5-4)$$

and the weighting factor matrices are for state and control variables are

$$\Phi = \begin{bmatrix} \phi_n & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \phi_{c1} & & & & & & 0 \\ \cdot & & \phi_{c2} & & & & & \cdot \\ \cdot & & & \phi_{c3} & & & & \cdot \\ \cdot & & & & \phi_{c4} & & & \cdot \\ \cdot & & & & & \phi_{c5} & & \cdot \\ 0 & \cdot & \cdot & \cdot & & & & \phi_{c6} \end{bmatrix} \quad (5-5)$$

with  $\Psi = 1$  (5-6)

The matrix of feedback gains is symmetric and written

$$K = K^T = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} & k_{15} & k_{16} & k_{17} \\ \cdot & k_{22} & k_{23} & k_{24} & k_{25} & k_{26} & k_{27} \\ \cdot & \cdot & k_{33} & k_{34} & k_{35} & k_{36} & k_{37} \\ \cdot & \cdot & \cdot & k_{44} & k_{45} & k_{46} & k_{47} \\ \cdot & \cdot & \cdot & \cdot & k_{55} & k_{56} & k_{57} \\ \cdot & \cdot & \cdot & \cdot & \cdot & k_{66} & k_{67} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & k_{77} \end{bmatrix} \quad (5-7)$$

The Riccati matrix equation is here repeated:

$$\dot{K} = -A^T K - KA + KB\Psi^{-1}B^T K - \Phi \quad (5-8)$$

For the infinite-interval problem the steady state solution of equation (5-8) is required. Thus a non-linear algebraic equation in the parameters  $k_{ij}$  must be solved. The feedback gains are seen to be sensitive to the specific

reactor parameters  $\beta$ ,  $\beta_1$ ,  $\lambda_1$  and  $\ell$ . The most sensitive of these parameters is the assumed mean neutron lifetime of the reactor system,  $\ell$ . The parameters  $\beta$ ,  $\beta_1$  and  $\lambda_1$  are functions of the type of fuel utilized and overall weight that delayed neutrons have in the neutron generation cycle. The numerical values of these variables are generally known fairly accurately and errors in the solution due to these variables are never great. The neutron lifetime  $\ell$ , on the other hand, is difficult to determine for dynamic operation. Zero-power-reactor frequency-domain studies are generally used to determine this quantity, but if many non-linear feedback effects are present, the value of  $\ell$  may be in error by a large factor. If  $\ell$  is not known accurately, however, statistical control theory should be used and a basic barrier to the problem is encountered.

The feedback gains are seen to be independent of the nominal states of the system. Thus, once the steady state solution of equation (5-8) is computed for a given set of reactor parameters, a feedback controller built of constant gain elements is sufficient for all operation no matter what nominal control and state variables are required. The optimal feedback control equation is

$$u = u^* - k_{11}(n-n^*) - \sum_{i=1}^6 k_{1(i+1)}(c_i - c_i^*) \quad (5-9)$$

and the optimal feedback reactivity is

$$\rho = \frac{\ell}{n} \left[ u^* - k_{11}(n-n^*) - \sum_{i=1}^6 k_{1(i+1)}(c_i - c_i^*) \right] \quad (5-10)$$

For adequate performance the selection of weighting factors is very important for this seventh-order process. These must be determined from the performance requirements. Equation (5-10) indicates that an auxiliary device is required to convert the pseudo-control to a reactivity control. The schematic of the optimal feedback control system is that given in figure 4 of chapter 3.

An indication of how sensitive the feedback gains of the dynamic process are with respect to the mean neutron lifetime  $\ell$ , can be obtained from a one group approximation of the six group model. The approximate dynamic process is

$$\dot{n} = u - \frac{\beta n}{\ell} + \lambda c \quad (5-11)$$

$$\dot{c} = \frac{\beta n}{\ell} - \lambda c \quad (5-12)$$

The abbreviated error criterion is

$$e(t) = \int_t^{\infty} \left\{ \phi_n [n(\tau) - n^*(\tau)]^2 + \phi_c [c(\tau) - c^*(\tau)]^2 + [u(\tau) - u^*(\tau)]^2 \right\} d\tau \quad (5-13)$$

The system matrix and the input matrix are respectively

$$A = \begin{bmatrix} -\beta/\ell & \lambda \\ \beta/\ell & -\lambda \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (5-14)$$

The resulting gain equations that must be solved are

$$\dot{k}_{11} = 2\frac{\beta}{\ell} (k_{11} - k_{12}) + k_{11}^2 - \phi_n \quad (5-15)$$

$$\dot{k}_{12} = \frac{\beta}{\ell} (k_{12} - k_{22}) + \lambda(k_{12} - k_{11}) + k_{11}k_{12} \quad (5-16)$$

$$\dot{k}_{22} = 2\lambda (k_{22} - k_{12}) + k_{12}^2 - \phi_c \quad (5-17)$$

where the  $k_{ij}$  are typical elements of the symmetric matrix

$$K = K^T = \begin{bmatrix} k_{11} & k_{12} \\ k_{12} & k_{22} \end{bmatrix} \quad (5-18)$$

The steady state solution of equations (5-15) through (5-17) is obtainable analytically. Only  $k_{11}$  and  $k_{12}$  are required for the optimal control equation, which is

$$u(\tau) = u^*(\tau) - k_{11}[n(\tau) - n^*(\tau)] - k_{12}[c(\tau) - c^*(\tau)] \quad (5-19)$$

The solutions for these gains, with  $\phi_c = 0$ , are

$$k_{11} = -\left[\frac{\beta}{\ell} + \lambda\right] + \sqrt{\left(\frac{\beta}{\ell} + \lambda\right)^2 + 2\lambda\sqrt{\phi_n} + \phi_n} \quad (5-20)$$

$$\text{and } k_{12} = \left(\lambda + \frac{\beta\lambda^2}{\ell}\right) + \frac{\ell\lambda}{\beta} \left\{ \sqrt{\phi_n} - \left[\left(\frac{\beta}{\ell} + \lambda\right)^2 + 2\lambda\sqrt{\phi_n} + \phi_n\right]^{\frac{1}{2}} \right\} \quad (5-21)$$

Since the gains are analytical functions of the generation time, a quantity that represents the effect of perturbations in  $\ell$  can be introduced here. This function is

termed a sensitivity function and is defined by the equation

$$S_y^x(z) = \frac{\delta y(z)/y(z)}{\delta x/x} \quad (5-22)$$

In this case  $y(z)$  corresponds to  $k_{11}(\ell)$ ,  $x$  corresponds to  $\ell$ , and  $\delta x$  corresponds to  $\delta \ell$ . Thus

$$y = k_{11}, \quad z = \ell \quad \text{and} \quad x = \ell. \quad (5-23)$$

The sensitivity function associated with  $k_{11}$  is

$$S_{k_{11}}^{\ell}(\ell) = \left\{ \left( 1 + \frac{\beta \lambda}{\ell} \right)^2 + \ell^2 [2\lambda \sqrt{\phi_n} + \phi_n] / \beta^2 \right\}^{-\frac{1}{2}} \quad (5-24)$$

This function expresses information concerning the changes in stability of the control system due to the non-exact determination of the mean neutron lifetime. It also indicates in what way errors are introduced if  $\ell$  is a variable parameter during the dynamic operation. Figure 15 illustrates the behavior of the sensitivity function for several values of the weighting factor  $\phi_n$ . The values of  $\beta$  and  $\lambda$  were arbitrarily set equal

$$\beta = 0.01 \quad \text{and} \quad \lambda = 0.1 \text{ sec}^{-1} \quad (5-25)$$

Several qualitative features are indicated in this figure. For values of  $\ell$  in the range typical of most reactors, namely  $10^{-9} < \ell < 10^{-3}$  seconds, the gain is a very sensitive function of neutron lifetime. It is easily seen from equation (5-20) that as  $\ell \rightarrow 0$ ,  $k_{11} \rightarrow \infty$ , and figure 15

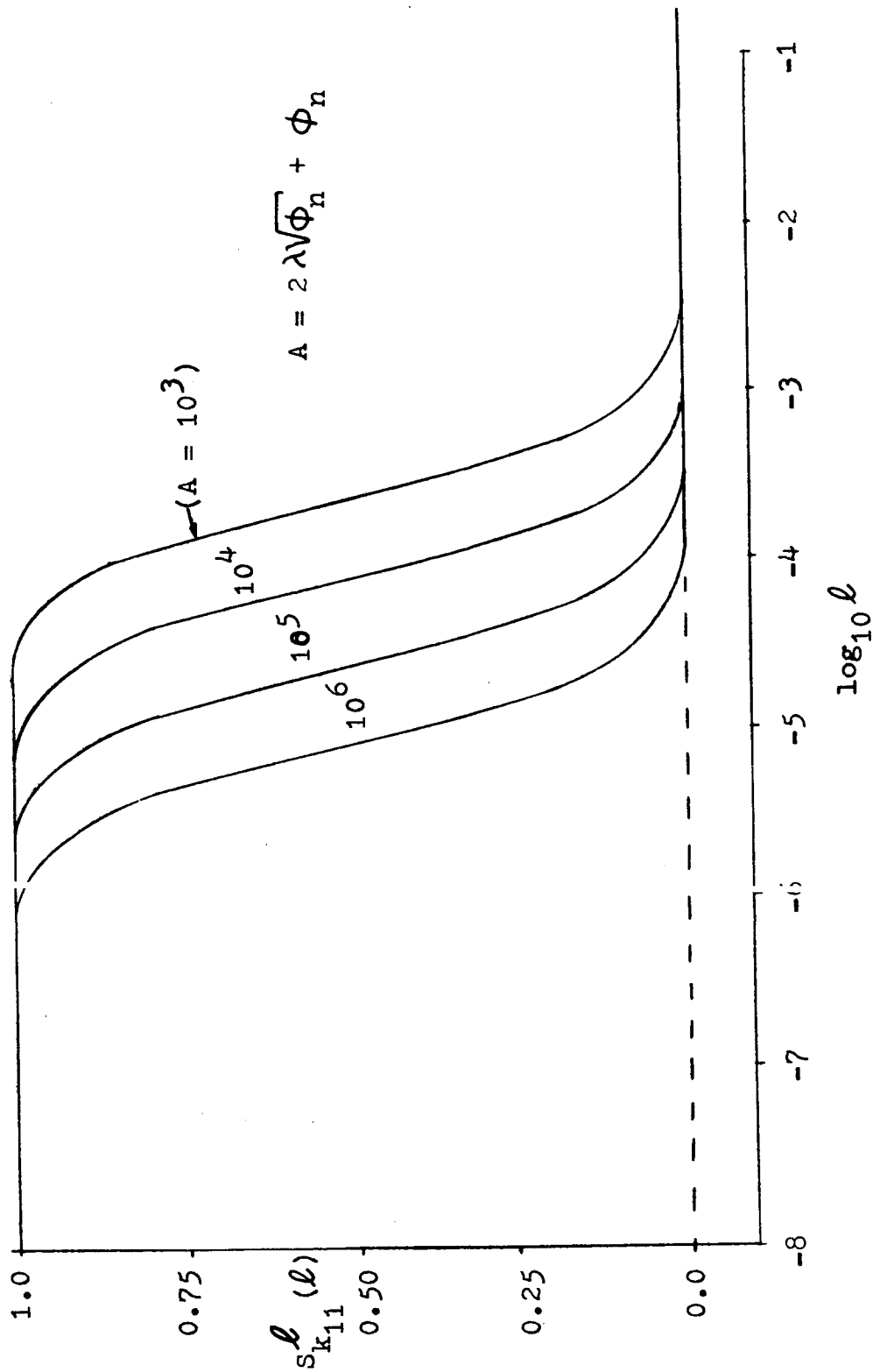


Figure 15  
Sensitivity of Feedback Gain,  $k_{11}$ , due to  $\ell$



indicates that, depending on the value of  $\phi_n$ , this may fall in the region of prime importance in nuclear reactor control. From this figure, a very good idea of what  $\phi_n$  should be to reduce gain sensitivity may be obtained. Qualitatively, the greater the value of  $\phi_n$  (or A on the figure) the less sensitive  $k_{11}$  is to perturbations in  $\ell$ . Zero sensitivity is desirable since this indicates that the gain  $k_{11}$  is approximately constant with respect to  $\ell$ . The transition is seen to occur over approximately two decades of values of  $\ell$ , and  $\phi_n$  should be chosen such that the appropriate approximate mean neutron lifetime falls near the end-point of this transition where  $S \approx 0$ . The gain must not, however, exceed physical limitations of the controller and this will place a constraint on the weighting factor  $\phi_n$ . The smallest  $\phi_n$  which most nearly satisfies these considerations should be used.

### Synthesis of Reactor Power Transfer

The problem of optimal power transfer of a TRIGA type nuclear reactor was studied by Rosztoczy (11, 12). In his study the optimal state and control variables were determined so as to minimize control rod energy while transferring the reactor power level from 10 kilowatts to 50 kilowatts in 0.47 seconds.

A bare thermal reactor, with an intrinsic feedback proportional to reactor power and one group of delayed neutrons, was used. The reactor kinetics are described by

the equations

$$\dot{n} = \frac{\rho n}{\ell} - \frac{\alpha n^2}{\ell} - \frac{\beta n}{\ell} + \lambda c \quad (5-26)$$

$$\dot{c} = \frac{\beta n}{\ell} - \lambda c \quad (5-27)$$

where  $\alpha$  is greater than zero, and  $\alpha n$  represents the power-feedback reactivity. Prior to the transfer process, the reactor is assumed to be in the steady state with the initial conditions

$$n(0) = n_0 = 10\text{kw}, \quad c(0) = c_0 = \frac{\beta n_0}{\ell \lambda} \quad (5-28)$$

The problem is to increase the power level to 50 kilowatts while minimizing

$$e(t) = \int_0^T [\dot{\rho}(\tau)]^2 d\tau \quad (5-29)$$

The rate of change of reactivity,  $\dot{\rho}$ , is considered to be a control variable and  $\rho$  is considered a state variable in this work. Thus the dynamic process is described by equations (5-26) and (5-27) together with equation (5-30):

$$\dot{\rho} = u \quad (5-30)$$

Figures 16, 17, and 18 show the open-loop optimal (or nominal) state variables,  $\rho^*$  and  $n^*$ , and the open-loop control variable,  $\dot{\rho}^*$ , for this problem. Constant acceleration control is required for minimum energy.

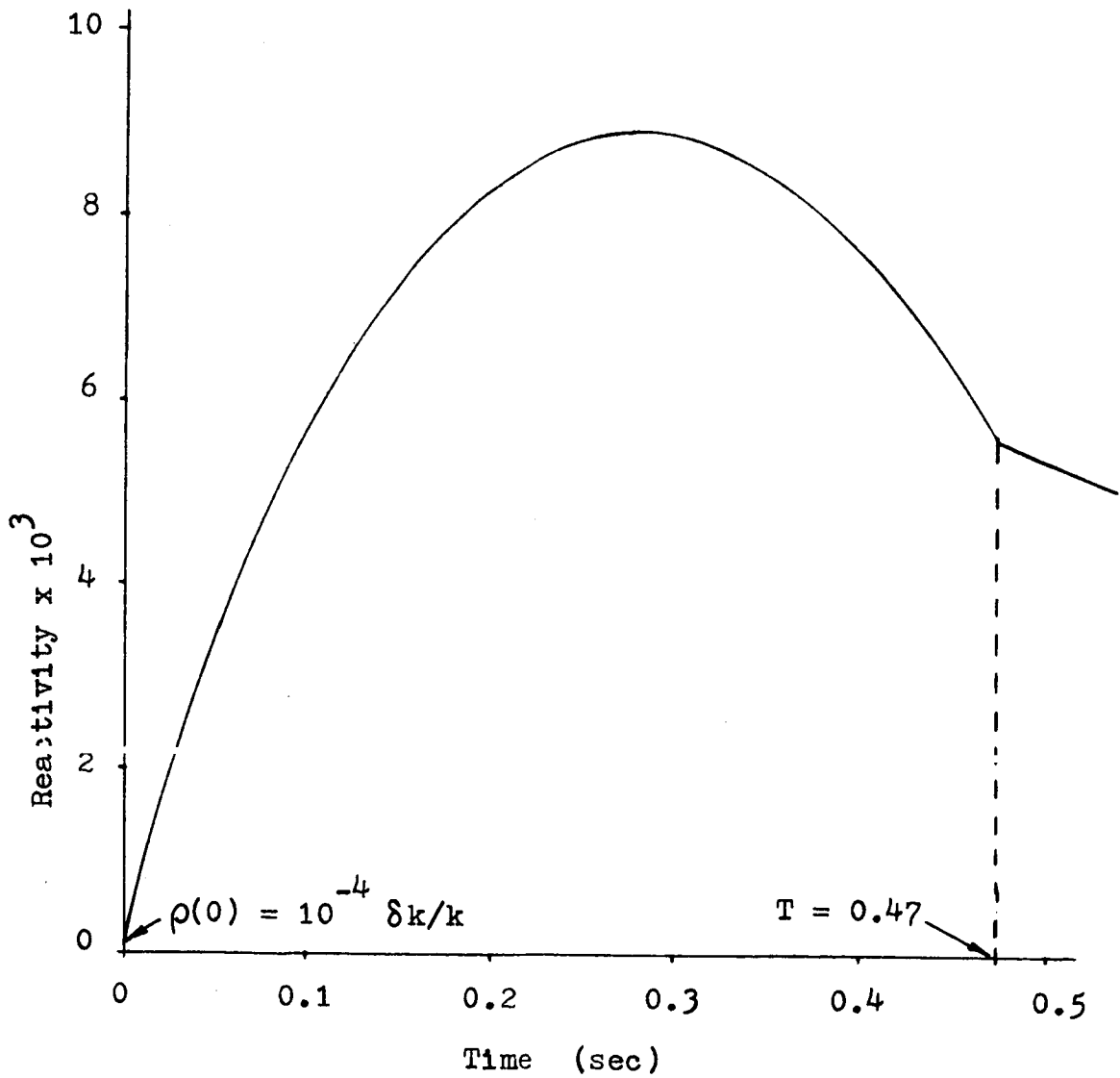


Figure 16  
Open-Loop Nominal Reactivity

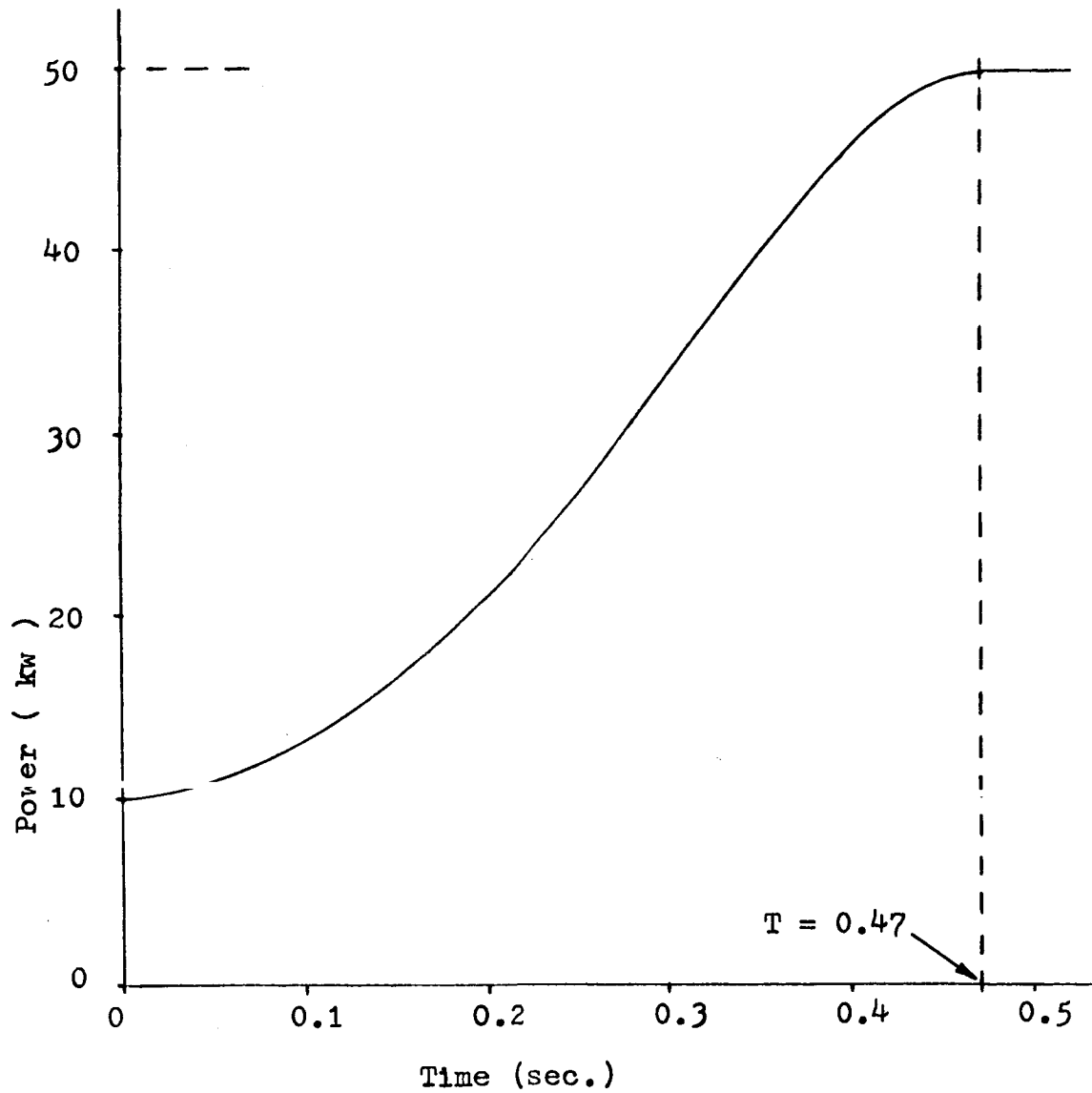


Figure 17  
Open-Loop Nominal Power

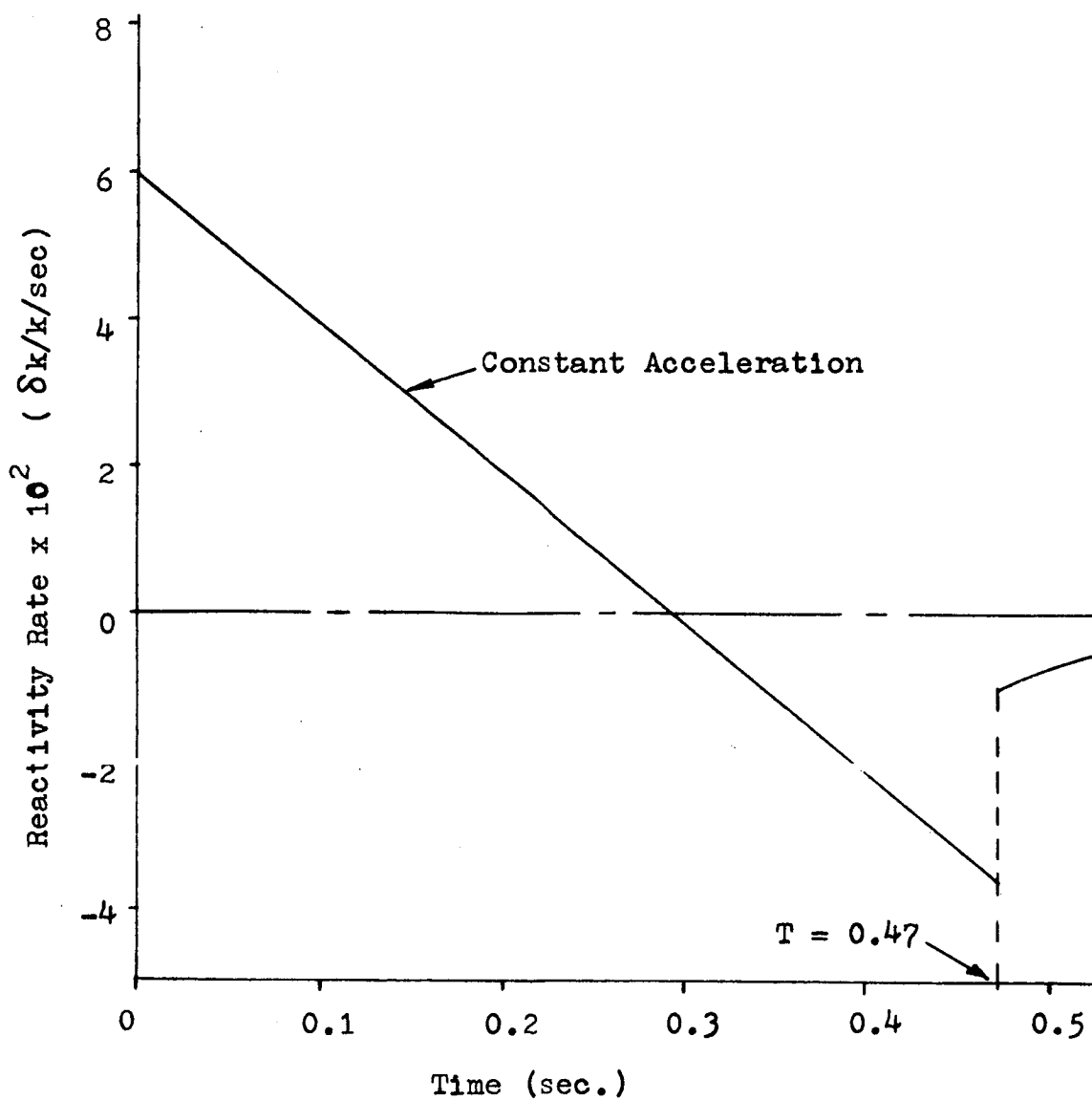


Figure 18  
Open-Loop Nominal Control

To synthesize this problem a closed-loop controller is determined to minimize the quadratic error criterion of equation (5-31).

$$e(0) = \int_0^{T=0.47} \left\{ \phi_n [n(\tau) - n^*(\tau)]^2 + \phi_\rho [\rho(\tau) - \rho^*(\tau)]^2 + [u(\tau) - u^*(\tau)]^2 \right\} d\tau \quad (5-31)$$

Note that no effort is made to weight precursor density errors in this problem. This does not mean, however, that there is no need to measure this state variable. This only indicates that no limit is placed on precursor density errors.

In this study a quasi-optimal feedback controller is found using the maximum principle. The linearized model describing perturbations about the nominal trajectories is represented

$$\delta \dot{\underline{x}}(\tau) = \begin{bmatrix} f_{11}(\tau) & \lambda & f_{13}(\tau) \\ \beta/l & -\lambda & 0 \\ 0 & 0 & 0 \end{bmatrix} \delta \underline{x}(\tau) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \delta \underline{u}(\tau) \quad (5-32)$$

$$\text{where } \delta \underline{x}(\tau) = [\underline{x}(\tau) - \underline{x}^*(\tau)] = \begin{bmatrix} n(\tau) - n^*(\tau) \\ c(\tau) - c^*(\tau) \\ \rho(\tau) - \rho^*(\tau) \end{bmatrix} \quad (5-33)$$

$$\text{and } \delta \underline{u}(\tau) = u(\tau) - u^*(\tau) \quad (5-34)$$

Also,

$$f_{11}(\tau) = [\rho^*(\tau) - 2\alpha n^*(\tau) - \beta] / l \quad (5-35)$$

$$\text{and } f_{13}(\tau) = n^*(\tau)/\ell \quad (5-36)$$

The Riccati matrix equation (3-82) must be solved for this problem. The time-varying gain matrix  $K(\tau)$  is

$$K = K^T = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ \cdot & k_{22} & k_{23} \\ \cdot & \cdot & k_{33} \end{bmatrix} \quad (3-37)$$

The resulting quasi-optimal feedback control is

$$\begin{aligned} u(\tau) = u^*(\tau) - k_{13}[n(\tau) - n^*(\tau)] - k_{23}[c(\tau) - c^*(\tau)] \\ - k_{33}[\rho(\tau) - \rho^*(\tau)] \end{aligned} \quad (5-38)$$

The differential equations that result from the expansion of the Riccati matrix equation (3-82) are found to be

$$\dot{k}_{11} = k_{13}^2 - 2[f_{11}k_{11} + \beta k_{12}/\ell] - \phi_n \quad (5-39)$$

$$\dot{k}_{12} = k_{13}k_{23} - \lambda k_{11} + \lambda k_{12} - f_{11}k_{12} - \beta k_{22}/\ell \quad (5-40)$$

$$\dot{k}_{13} = k_{13}k_{33} - f_{13}k_{11} - f_{11}k_{13} - \beta k_{23}/\ell \quad (5-41)$$

$$\dot{k}_{22} = k_{23}^2 - 2\lambda k_{12} + 2\lambda k_{22} \quad (5-42)$$

$$\dot{k}_{23} = k_{23}k_{33} - f_{13}k_{12} - \lambda k_{13} + \lambda k_{23} \quad (5-43)$$

$$\dot{k}_{33} = k_{33}^2 - 2f_{13}k_{13} - \phi_\rho \quad (5-44)$$

The boundary conditions for equations (5-39) through (5-44) are given by the relation

$$K(T=0.47) = [0] \quad (5-45)$$

where all elements of  $[0]$  are zero.

Figure 19 illustrates the (time-varying) feedback gain programs of  $k_{13}$ ,  $k_{23}$ , and  $k_{33}$  as a function of time for weighting factors of

$$\begin{aligned} \phi_n &= 10^{-6} \text{ sec}^{-2} \text{kw}^{-2} \\ \phi_c &= 0 \\ \phi_p &= 25 \text{ sec}^{-2} \end{aligned} \quad (5-46)$$

and

$$\psi = 1$$

and for assumed reactor parameters of

$$\begin{aligned} \alpha &= 10^{-5} \text{ kw}^{-1} & \lambda &= 0.1 \text{ sec}^{-1} \\ \beta &= 0.0064 & n^*(0) &= 10 \text{kw} \\ \ell &= 10^{-3} \text{ sec} & n^*(T) &= 50 \text{kw} \\ T &= 0.47 \end{aligned} \quad (5-47)$$

The digital computer was used to get the solution of these gains (13).

In practice, time-varying feedback gains are not desirable. The synthesis problem of this example was simulated by means of an analog computer. Simulation of the time-varying gains on the analog computer can be accomplished either by direct solution of equations (5-39) through (5-44) along with the simulation process, or by reproducing these gains with function generators. The first method requires more computer components than were



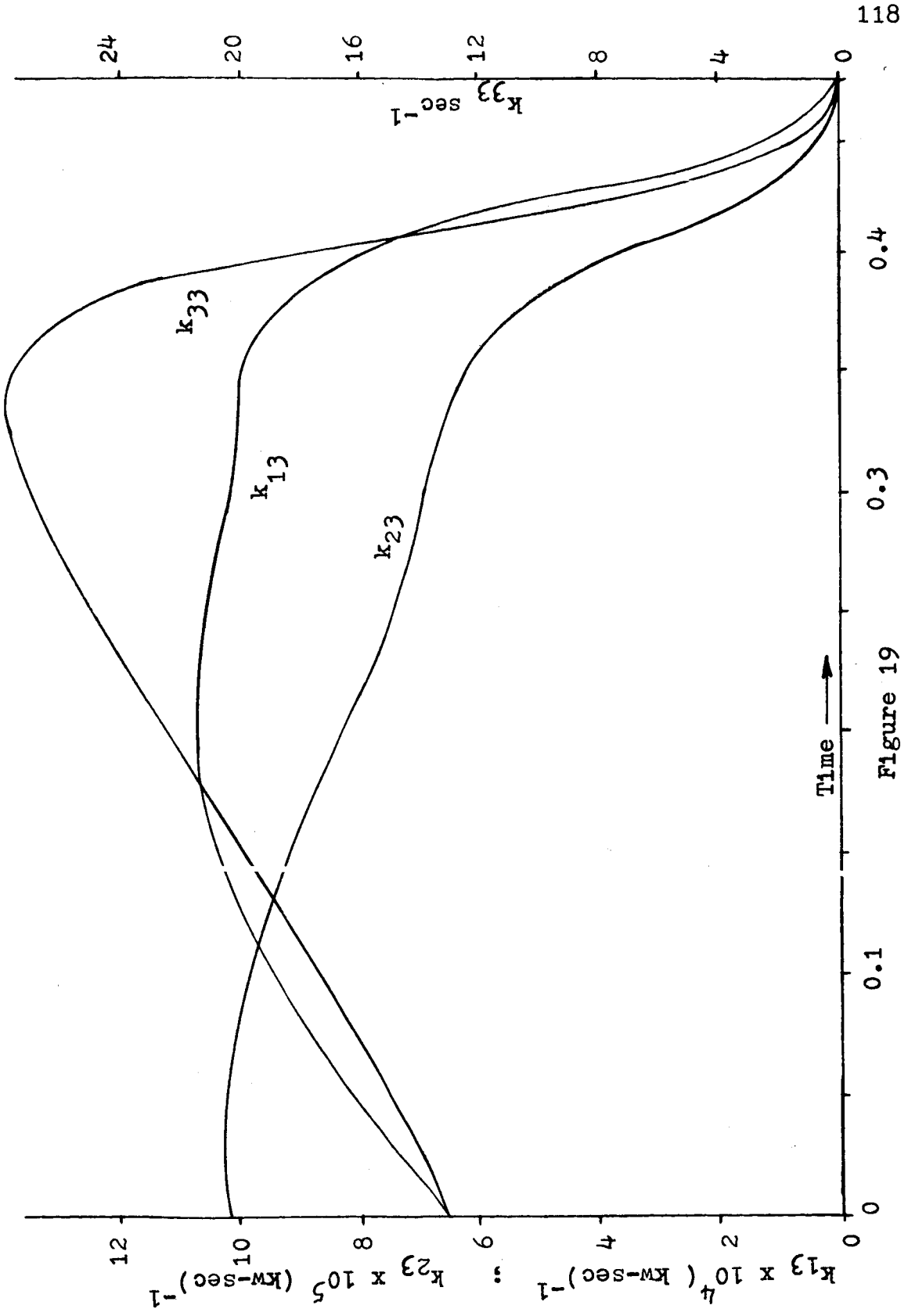


Figure 19

Time-Varying Gains for Reactor Power Transfer

available for this work. The latter method has the disadvantage that any changes in initial conditions or weighting factors require reprogramming of the function generators.

For the simulation study of this example, the time-varying gains of figure 19 were averaged over the control interval by the simple relation

$$\overline{k_{13}} = \frac{1}{0.47} \int_0^{0.47} k_{13}(\tau) d\tau \quad (i = 1, 2, 3) \quad (5-48)$$

These constant gains are easily programmed on the analog computer with coefficient potentiometers. This further quasi-optimal approximation permits easy simulation for many different average gain settings.

In analog computer synthesis the model and the "physical plant" are simulated simultaneously. Two nearly identical sets of dynamic equations are programmed; one generating the nominal control and state variables, the other contaminated with external noise. The contaminated system is driven by the nominal control program and, because of the noise, deviates from the nominal trajectories. The output error vector is computed, multiplied by the quasi-optimal feedback gains, and added to the nominal control variables. The result is a linear quasi-optimal feedback controller.

Figure 20 shows the results of this analog computer simulation. Three different conditions are illustrated for

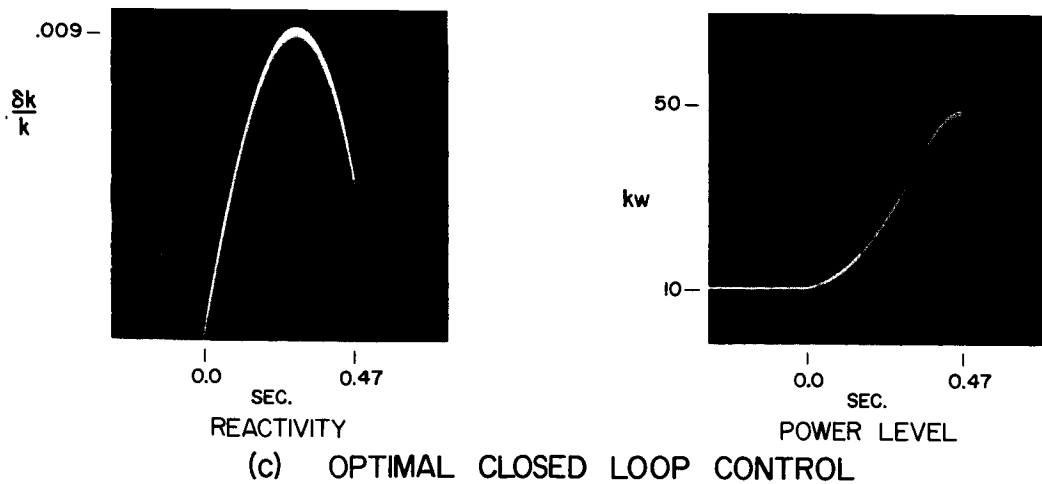
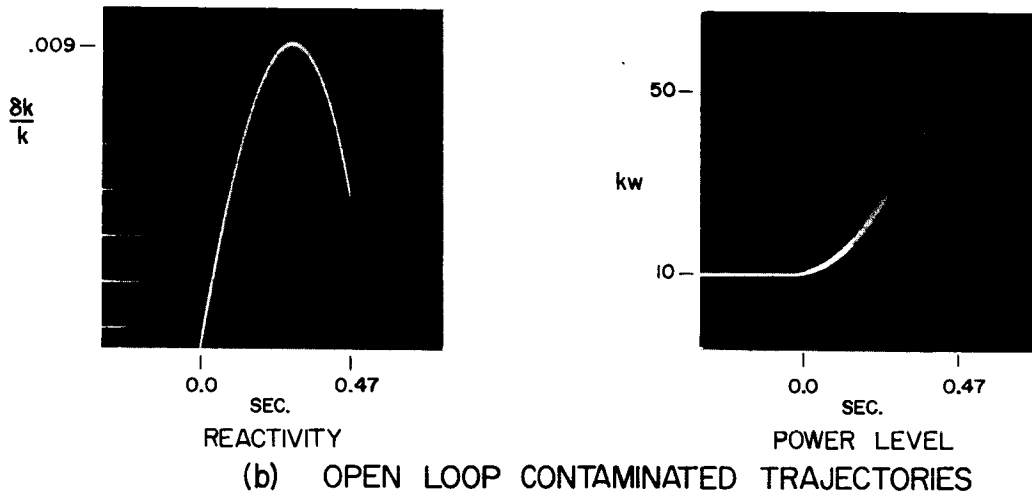
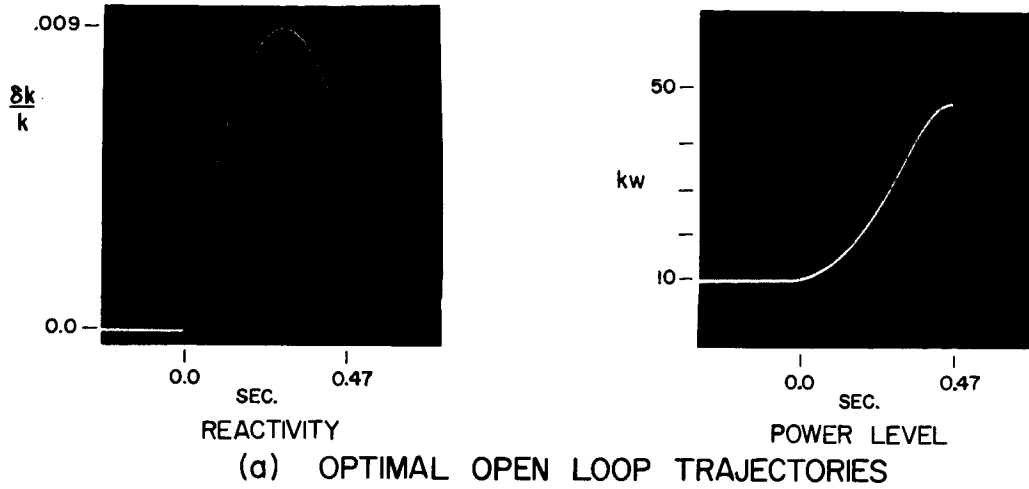


Figure 20

Reactivity and Power for Nuclear  
Reactor Power Transfer

the states of power level and reactivity. These conditions are:

- (a) nominal open-loop trajectories
  - (b) open-loop contaminated state trajectories
- and (c) the effect of closed-loop quasi-optimal control on the contaminated trajectories.

The closed-loop constant gain settings for figure 20 were

$$k_{13} = 1.1 \times 10^{-3} \text{ (kw-sec)}^{-1}$$

$$k_{23} = 7.1 \times 10^{-5} \text{ (kw-sec)}^{-1}$$

and  $k_{33} = 15.3 \text{ (sec}^{-1}\text{)}$

From figure 20(c) it is seen that constant gain feedback control can compensate for the noise to within one percent.

Constant gains were found to give adequate performance when the amplitude of the noise remained below a certain saturation level, where the system would become unstable. However, as these gains were increased, so was the point at which compensation was exceeded. The simplicity of using constant gains, in most instances, outweighs the difficulties of constructing a time-varying quasi-optimal controller.

#### Nuclear Rocket Engine Start-up Synthesis

The dynamic behavior of a nuclear rocket engine with a bleed turbo-pump or a topping turbo-pump propellant drive is approximately represented by the set of non-linear differential equations (8, 9)

$$\dot{n} = \frac{\rho_t - \beta}{\ell} n + \lambda c \quad (5-50)$$

$$\dot{c} = \frac{\beta n}{\ell} - \lambda c \quad (5-51)$$

and

$$\dot{T} = \frac{n}{M_c} - \frac{T}{\tau_T} \quad (5-52)$$

where the system state variables are  $n(t)$ ,  $c(t)$ , and  $T(t)$ ,  
and where

- $n$  = power
- $c$  = space-average (one group) precursor density
- $T$  = space-average core temperature
- $\rho_t$  = total reactivity
- $M_c$  = mean effective heat capacity of reactor core
- $\tau_T$  = heat-exchanger thermal time constant

Equation (5-52) represents the heat exchange equation which is coupled with the neutronics through core temperature and propellant flow rate in the form of reactivity. This total reactivity thus consists of the control rod reactivity  $u_1$ , a control variable, the propellant density reactivity  $\delta\rho_p$ , and the temperature reactivity  $\delta\rho_T$ :

$$\rho_t = u_1 + \delta\rho_T + \delta\rho_p \quad (5-53)$$

where  $\delta\rho_T = c_T T(t)$  (5-54)

and  $\delta\rho_p = c_p u_2(t) / \sqrt{T(t)}$  (5-55)

Here  $u_2(t)$  is the propellant mass flow rate, also a control variable. Thus two control variables are required for dynamic operation. Usually the constants in equations (5-54) and (5-55) are  $c_T < 0$  and  $c_p > 0$ . Finally, the heat-exchanger thermal "time constant" is not a constant, but depends on the propellant flow rate:

$$\tau_T = (bu_2)^{-1} \quad (5-56)$$

where  $b$  is a constant of proportionality at rated design flow rate.

For this reactor model, Mohler (8) has determined a set of nominal optimal open-loop state and control variables which minimize propellant consumption during the start-up operation. The reactor system is subject to the following constraints

$$\begin{aligned} u_a &\leq u_2(t) \leq u_b \\ n(t) &\leq n_{\max} \\ T(t) &\leq T_{\max} \end{aligned} \quad (5-57)$$

and

$$\dot{T} \leq \dot{T}_{\max} = \alpha$$

where  $\gamma$  is a positive number greater than 1.

Figure 21 shows the nominal state variable trajectories for core power and temperature with  $\gamma = 1.6$ . Figure 22 shows the nominal control variable trajectories for this case. The nominal control trajectories are seen to be

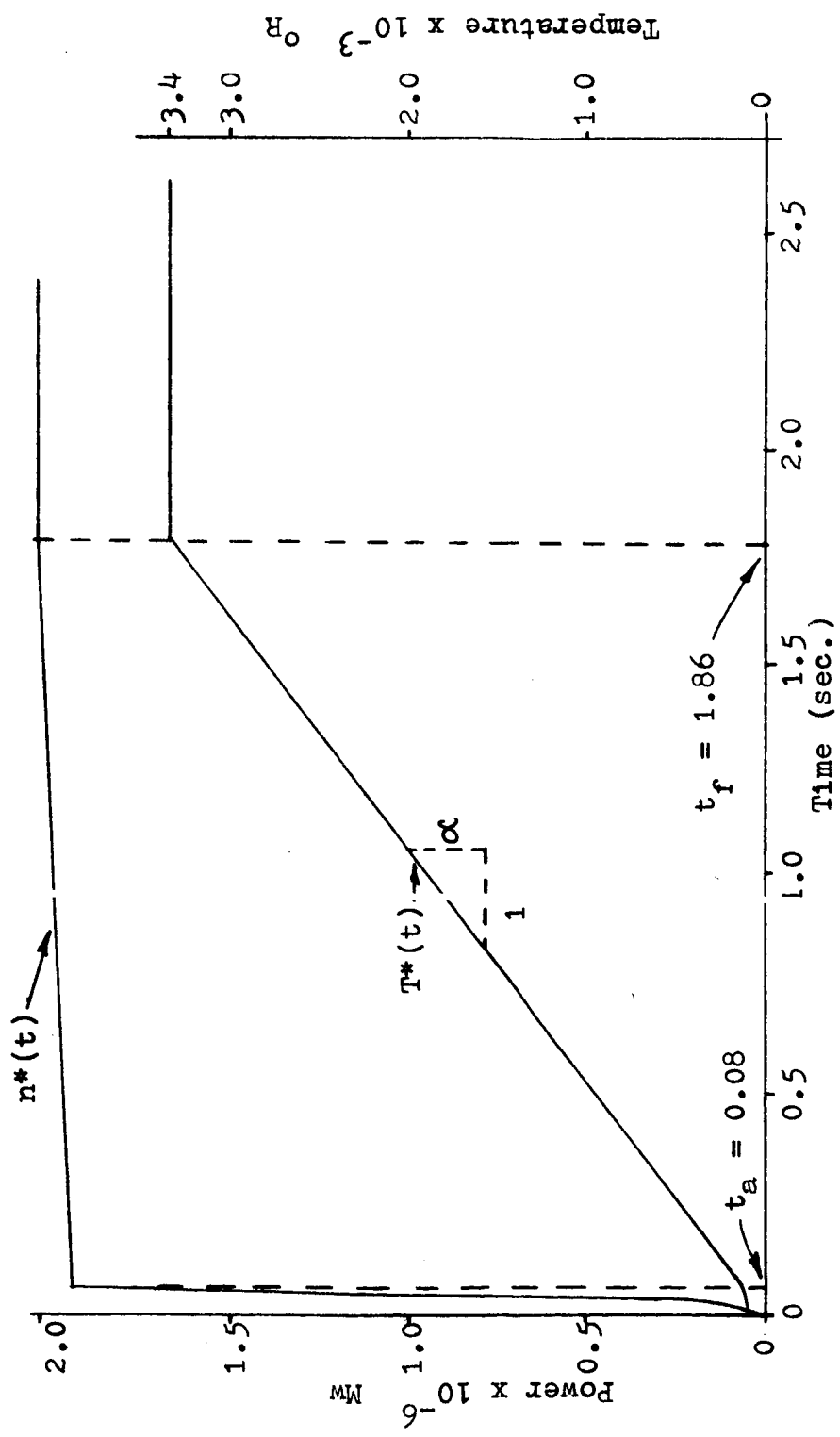


Figure 21  
Nominal State Trajectories;  $\gamma = 1.6$

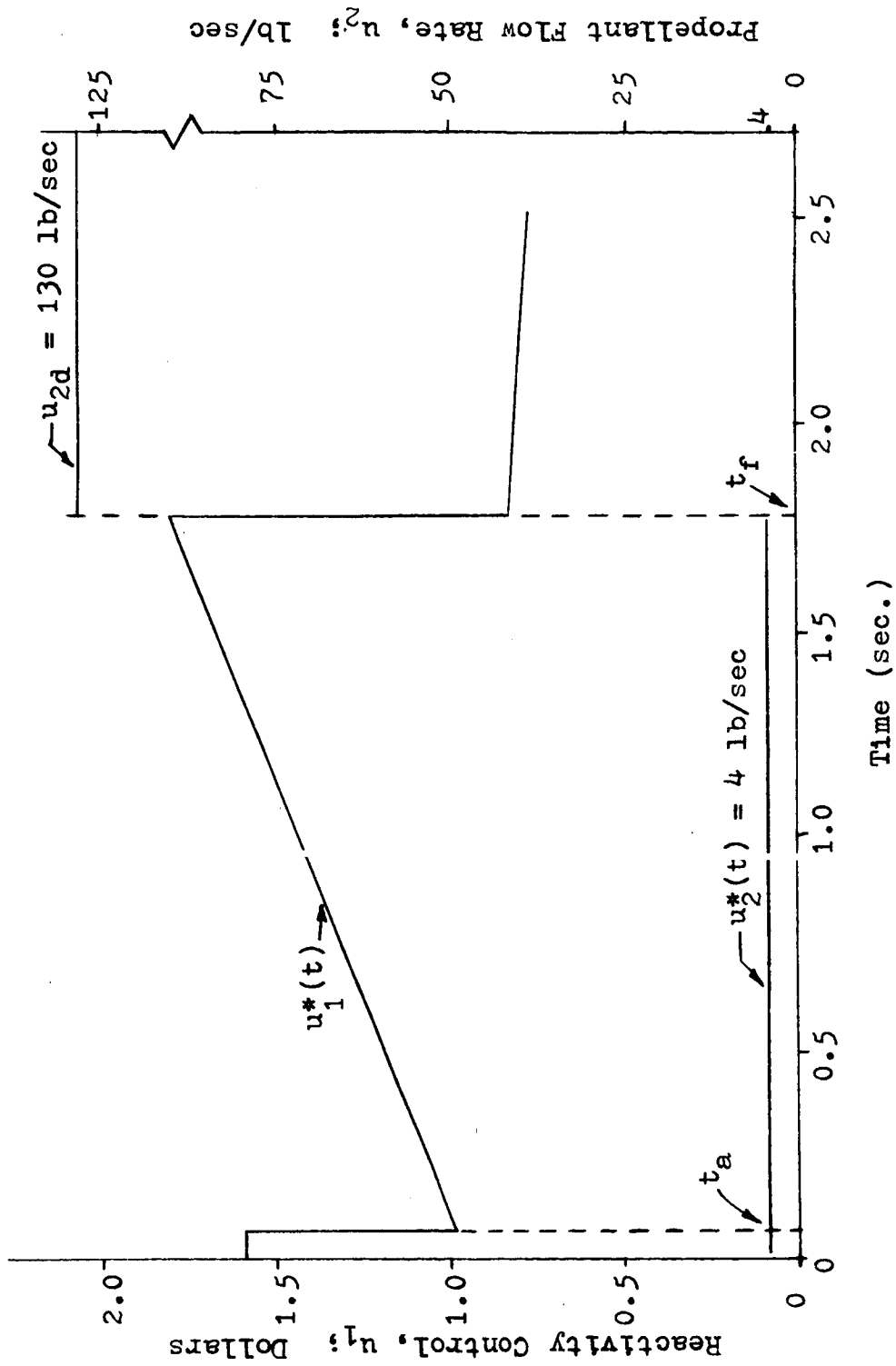


Figure 22

Nominal Control Variables;  $\gamma = 1.6$



discontinuous functions of time. Maximum total reactivity is initially applied in order to perform the start-up operation in a minimum time. The first switching time,  $t_a$ , occurs at the point where the rate of core temperature rise reaches its maximum value. At this time, control-rod reactivity is programmed such that  $\dot{T}$  remains at, but does not exceed, its maximum value. Beyond  $t_a$ , both reactor power and temperature increase linearly with time. Figure 21 indicates that both temperature and power reach their design values simultaneously at  $t_f$  while the propellant flow rate is maintained at its minimum value throughout the control period. Actually, if the nominal open-loop controls of figure 22 were used to drive the model of equations (5-50) through (5-52), core power would not reach its design value when temperature does. An additional switching time, say  $t_b$ , would occur very near the terminal time  $t_f$  where a short-duration control is applied to bring the power to its design value simultaneously with the design value of the temperature. However, a feedback controller designed to minimize errors in the state and control variables is used here to eliminate this switching time. Table I gives the hypothetical design parameters of this study.

As in all previous examples the desired optimal feedback controller minimizes errors in state trajectories and control motions. The error criterion is

TABLE I

## HYPOTHETICAL NUCLEAR ROCKET ENGINE PARAMETERS

Maximum Reactor Power, $n_{\max}$	2260 Mw
Minimum Propellant Flow Rate, $u_a$	4 lb/sec
Design Propellant Flow Rate, $u_{2d}$	130 lb/sec
Design Maximum Average Core Temperature, $T_{\max}$	3400 °R
Propellant Inlet Temperature, $T(0)$	120 °R
Mean Effective Neutron Lifetime, $l$	$3 \times 10^{-5}$ sec
Effective Delayed Neutron Fraction, $\beta$	0.0065
Design Propellant Reactivity, $\delta\rho_p(t_f)$	0.0065
Design Temperature Reactivity, $\delta\rho_T(t_f)$	-0.0065
Effective One-Group Delay Constant, $\lambda$	0.10 sec <sup>-1</sup>
Effective Core-Mass Heat Capacity, $M_c$	1140 Btu/°R
Maximum Core Temperature Rate of Rise, $\alpha$	1800 °R/sec
First Switching Time, $t_a$	0.08 sec
Terminal Switching Time, $t_f$	1.00 sec

$$e(0) = \int_0^{t_f} \left\{ \phi_n [n(\tau) - n^*(\tau)]^2 + \phi_T [T(\tau) - T^*(\tau)]^2 + \psi_1 [u_1(\tau) - u_1^*(\tau)]^2 + \psi_2 [u_2(\tau) - u_2^*(\tau)]^2 \right\} d\tau \quad (5-58)$$

Here again, errors in precursor density are not considered.

A quasi-optimal feedback controller based on the maximum principle is again evaluated. The linearized dynamic process describing perturbations of state and control variables is represented by

$$\delta \dot{\underline{x}} = \begin{bmatrix} a_{11}(\tau) & \lambda & a_{13}(\tau) \\ \beta/\ell & -\lambda & 0 \\ 1/M_c & 0 & a_{33}(\tau) \end{bmatrix} \delta \underline{x} + \begin{bmatrix} b_{11}(\tau) & b_{12}(\tau) \\ 0 & 0 \\ 0 & b_{32}(\tau) \end{bmatrix} \delta \underline{u} \quad (5-59)$$

where

$$\delta \underline{x} = \begin{bmatrix} \delta n \\ \delta c \\ \delta T \end{bmatrix} \quad \text{and} \quad \delta \underline{u} = \begin{bmatrix} \delta u_1 \\ \delta u_2 \end{bmatrix} \quad (5-60)$$

$$\text{and} \quad a_{11}(\tau) = \left[ u_1^* + c_T T^* + \frac{c_p u_2^* - \beta}{\sqrt{T^*}} \right] / \ell \quad (5-61)$$

$$a_{13}(\tau) = \left[ c_T n^* - \frac{c_p u_2^* n^*}{2\sqrt{T^* \beta}} \right] / \ell \quad (5-62)$$

$$a_{33}(\tau) = - b u_2^* \quad (5-63)$$

$$b_{11}(\tau) = n^* / \ell \quad (5-64)$$

$$b_{12}(\tau) = c_p n^* / \ell \sqrt{T^*} \quad (5-65)$$

$$b_{32}(\tau) = - b T^* \quad (5-66)$$

The Riccati-type differential equations that must be solved to determine the time-varying feedback gains are

$$\begin{aligned} \dot{k}_{11} = & Ak_{11}^2 - 2Bk_{11}k_{13} + Ck_{13}^2 - 2a_{11}k_{11} - 2\beta k_{12}/\ell \\ & - 2k_{13}/M_c - \phi_n \end{aligned} \quad (5-67)$$

$$\begin{aligned} \dot{k}_{12} = & Ak_{11}k_{12} - Bk_{11}k_{23} - Bk_{12}k_{13} + Ck_{13}k_{23} - \lambda k_{11} + \\ & (\lambda - a_{11})k_{12} - \beta k_{22}/\ell - k_{23}/M_c \end{aligned} \quad (5-68)$$

$$\begin{aligned} \dot{k}_{13} = & Ak_{11}k_{13} - Bk_{11}k_{33} - Bk_{13}^2 + Ck_{13}k_{33} - a_{13}k_{11} \\ & - (a_{33} + a_{11})k_{13} - \beta k_{23}/\ell - k_{33}/M_c \end{aligned} \quad (5-69)$$

$$\dot{k}_{22} = Ak_{12}^2 - 2Bk_{12}k_{23} + Ck_{23}^2 - 2\lambda k_{12} + 2\lambda k_{22} \quad (5-70)$$

$$\begin{aligned} \dot{k}_{23} = & Ak_{12}k_{13} - Bk_{11}k_{33} - Bk_{13}k_{23} + Ck_{23}k_{33} - a_{13}k_{12} \\ & - \lambda k_{13} - (a_{33} - \lambda)k_{23} \end{aligned} \quad (5-71)$$

$$\begin{aligned} \dot{k}_{33} = & Ak_{13}^2 - 2Bk_{13}k_{33} + Ck_{33}^2 - 2a_{13}k_{13} - 2a_{33}k_{33} - \phi_T \end{aligned} \quad (5-72)$$

where

$$A(\tau) = \frac{b_{11}^2}{\psi_1} + \frac{b_{12}^2}{\psi_2} \quad (5-73)$$

$$B(\tau) = - \frac{b_{12}b_{32}}{\psi_2} \quad (5-74)$$

and 
$$C(\tau) = \frac{b_{32}^2}{\psi_2} \quad (5-75)$$

The equations that describe the quasi-optimal feedback control variables are

$$u_1 = u_1^* - \frac{b_{11}}{\gamma_1} \left[ k_{11}(n-n^*) + k_{12}(c-c^*) + k_{13}(T-T^*) \right] \quad (5-76)$$

and

$$u_2 = u_2^* - \frac{b_{12}}{\gamma_2} \left[ k_{11}(n-n^*) + k_{12}(c-c^*) + k_{13}(T-T^*) \right] \\ - \frac{b_{32}}{\gamma_2} \left[ k_{13}(n-n^*) + k_{23}(c-c^*) + k_{33}(T-T^*) \right] \quad (5-77)$$

Figure 23 is a plot of the time-varying feedback gains associated with the nominal trajectories of power and temperature for the two control variables. The switching time at  $t_b < t_f$  is included and the value of  $\gamma = 1.60$  is used. The gains were obtained from digital computer solutions of equations (5-67) through (5-72) (13). The gains for the precursor density are not illustrated; they have the same general shape as those illustrated. For the gains in this figure, the following relations hold:

$$k_{1n} = k_{11}(\tau)b_{11}(\tau)/\gamma \quad (5-78)$$

$$k_{1T} = k_{13}(\tau)b_{11}(\tau)/\gamma \quad (5-79)$$

$$k_{2n} = \left[ k_{11}(\tau)b_{12}(\tau) + k_{13}(\tau)b_{32}(\tau) \right] / \gamma_2 \quad (5-80)$$

$$k_{2T} = \left[ k_{13}(\tau)b_{12}(\tau) + k_{33}(\tau)b_{32}(\tau) \right] / \gamma_2 \quad (5-81)$$

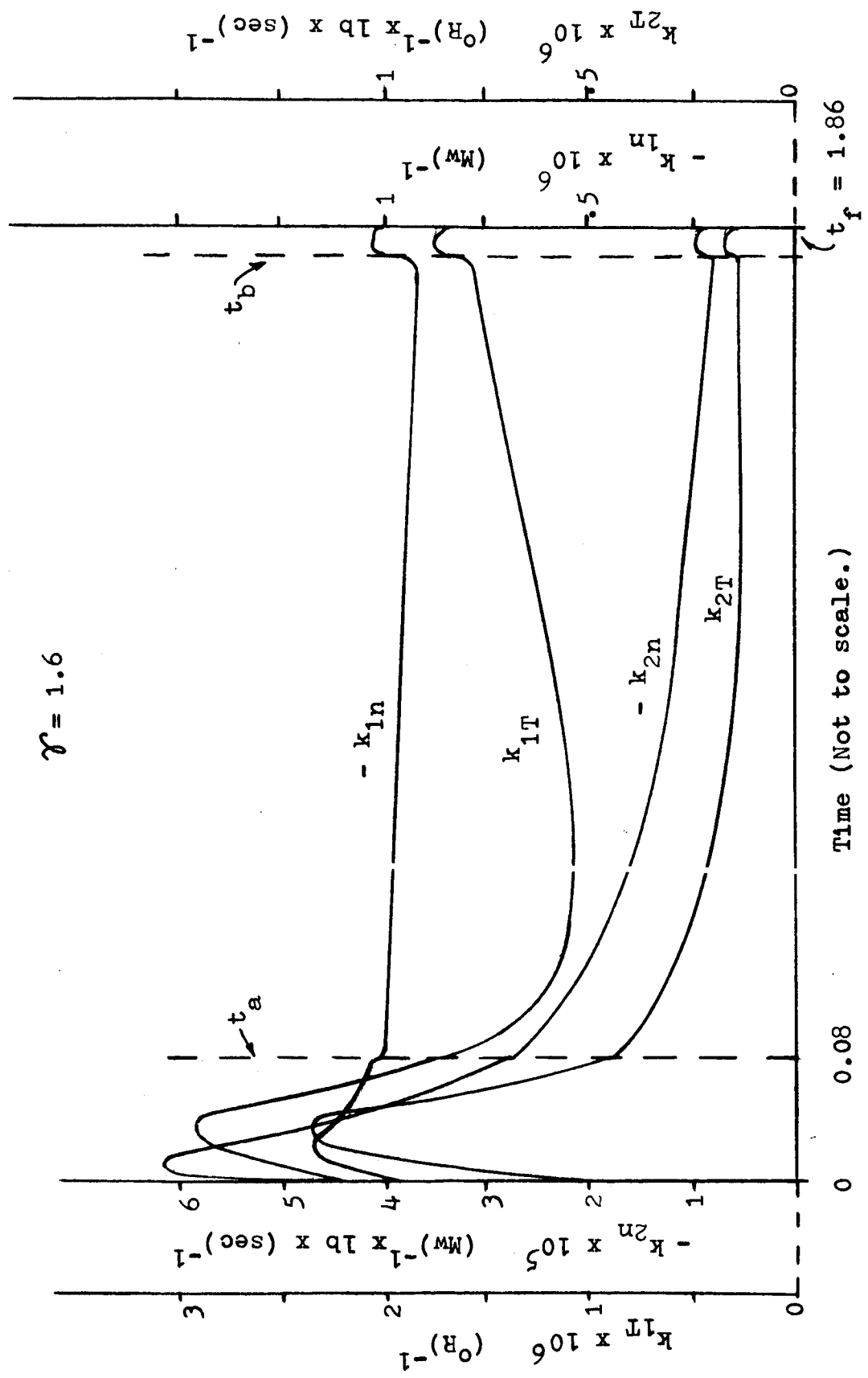


Figure 23

Time-Varying Feedback Gains for a Nuclear Rocket Engine

The weighting factors used to determine these gains were computed using the techniques presented in chapter 3.

They are:

$$\begin{aligned} \psi_1 &= 1 \\ \psi_2 &= 4 \times 10^{-4} \quad (\text{sec/lb})^2 \\ \phi_n &= 2.5 \times 10^{-10} \quad (\text{Mw}^{-2}) \\ \phi_T &= 10^{-10} \quad (^\circ\text{R})^{-2} \end{aligned} \tag{5-82}$$

Simulation of the rocket engine start-up problem using quasi-optimal feedback control synthesis was programmed for the analog computer (14). As in the previous example, time-averages of the gains of figure 23 were used to further simplify the study. The average gain settings for most of this study were

$$\begin{aligned} k_{1n} &= -1 \times 10^{-6} \quad (\text{Mw}^{-1}) \\ k_{1c} &= -3 \times 10^{-9} \quad (\text{Mw}^{-1}) \\ k_{1T} &= 2 \times 10^{-6} \quad (^\circ\text{R})^{-1} \\ k_{2n} &= -2 \times 10^{-5} \quad (\text{lb sec}^{-1} \text{ Mw}^{-1}) \\ k_{2c} &= -6 \times 10^{-8} \quad (\text{lb sec}^{-1} \text{ Mw}^{-1}) \\ k_{2T} &= 5 \times 10^{-6} \quad (\text{lb sec}^{-1} \text{ }^\circ\text{R}^{-1}) \end{aligned} \tag{5-83}$$

Here again, both the model and the "physical plant" were

simulated together. The entire feedback controller was constructed of analog components. In the actual tests of KIWI nuclear rocket engines and the ROVER engine the nominal variables, including controls, power, etc., are in fact simulated via the analog computer.

Figure 24 shows the contaminated open-loop control variables used in this study. Figure 25 is a plot of the resulting open-loop power trajectory and shows the effect of closed-loop feedback control for the gain settings of equation (5-83). In figure 26 the average core temperature for open-loop and two closed-loop conditions are shown. The feedback effect is obvious. The gain vectors  $\underline{k}_1$  are given by the equation

$$\underline{k}_1 = \begin{bmatrix} k_{1n} \\ k_{1c} \\ k_{1T} \\ \vdots \\ k_{2n} \\ k_{2c} \\ k_{2T} \end{bmatrix} \quad (i = 1, 2) \quad (5-84)$$

The figure illustrates that the greater the norm value of the vector  $\underline{k}_1$  the better the compensation. The norm is defined as

$$\|\underline{k}\| = (\underline{k}^T \underline{k})^{\frac{1}{2}} \quad (5-85)$$



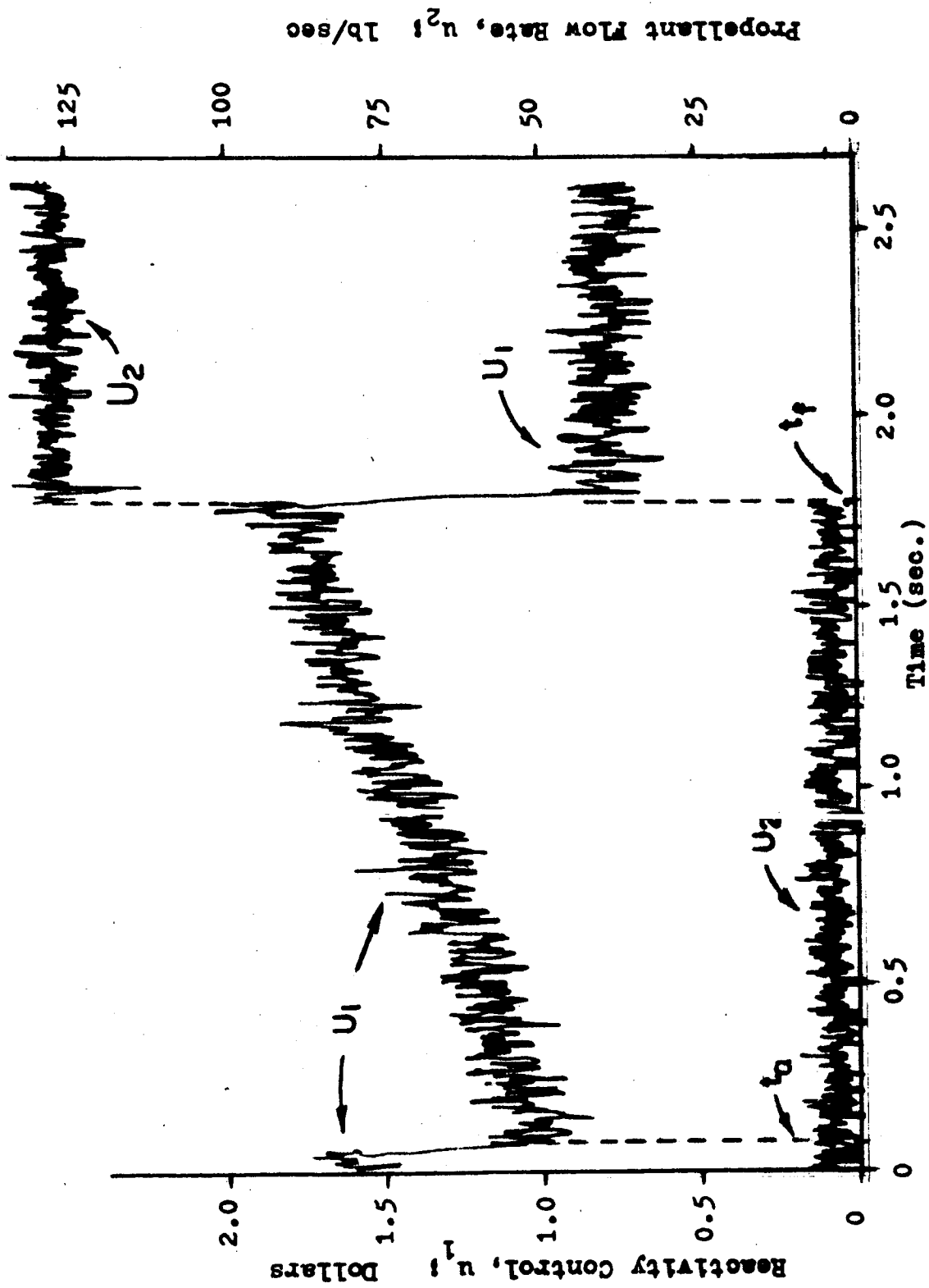


Figure 24  
 Contaminated Open-Loop Controls;  $\gamma = 1.6$

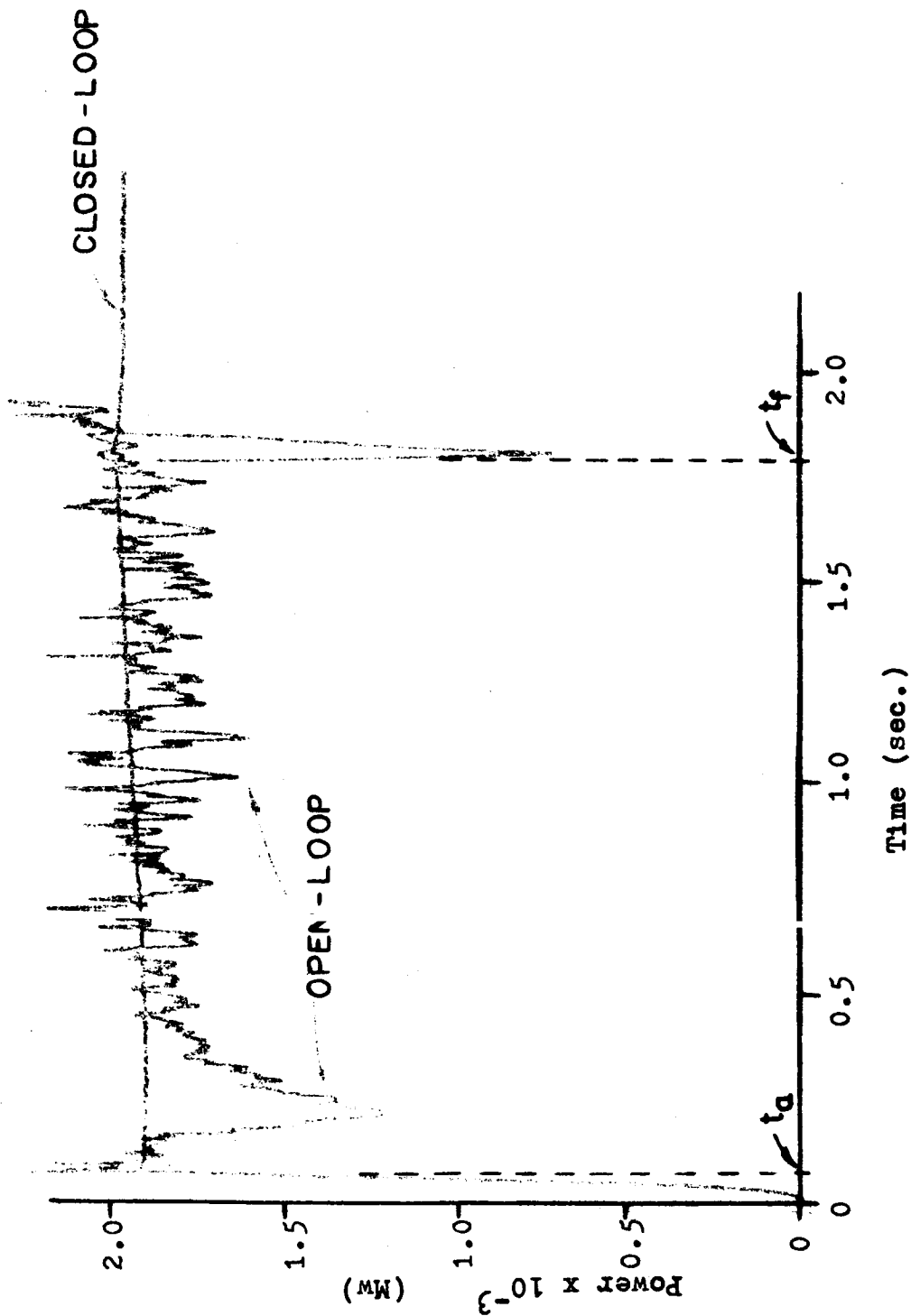


Figure 25

Effect of Feedback Control on Power

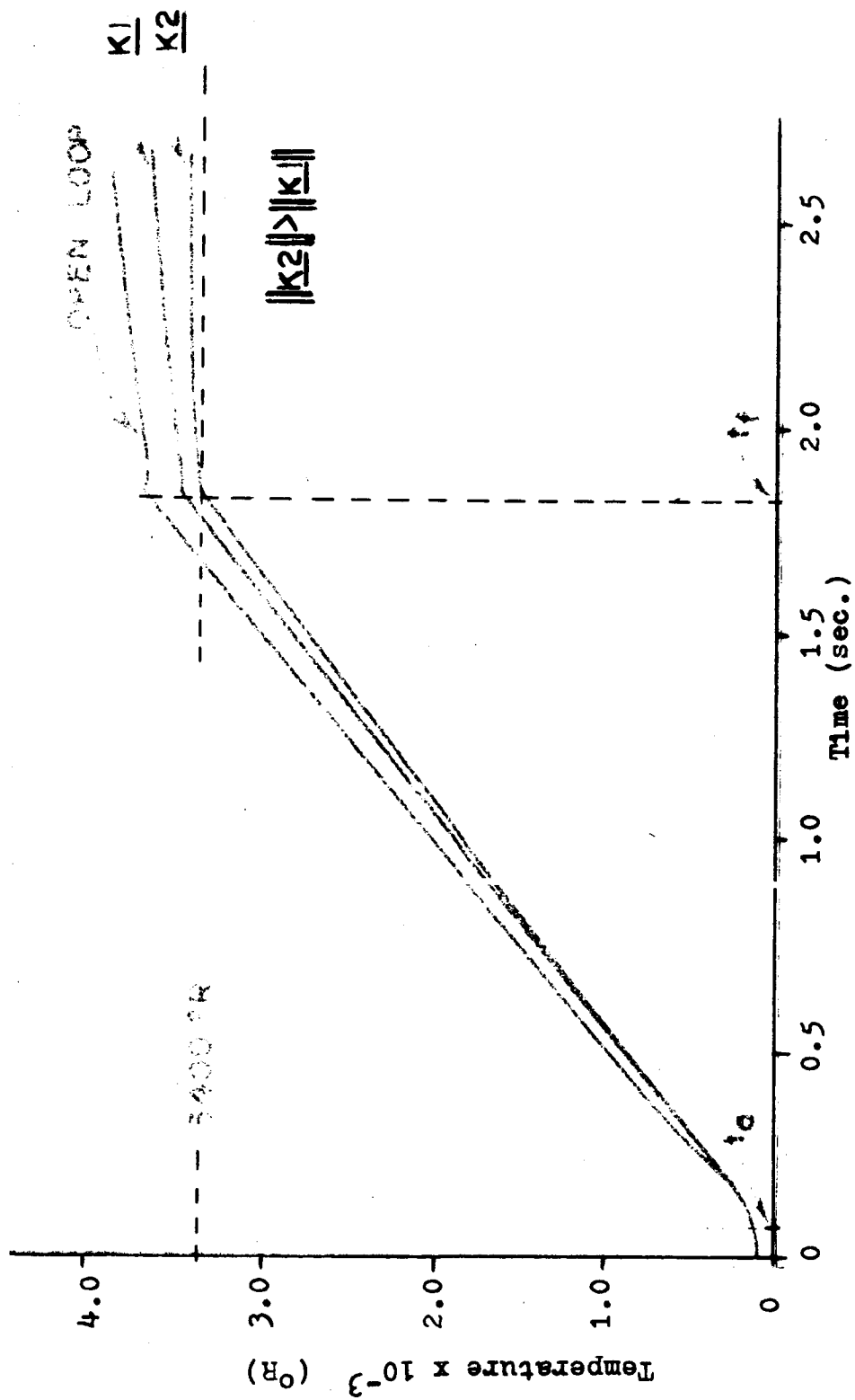


Figure 26  
Effect of Feedback Control on Temperature

and is representative of the length, or magnitude, of the vector in the six-dimensional gain space. The gain vector  $\underline{k}_2$  corresponds to the values of equation (5-83).

In this example, as evidenced from the figures, quasi-optimal control appears to give adequate compensation. The actual control synthesis problem, however, is non-linear and very much more complex than in this example. More extensive simulation is required to investigate other configurations. Dynamic programming might give a better performance controller. It is felt, however, that the usefulness of this technique is substantiated here.

## REFERENCES

1. Bellman, R. E., Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1957.
2. Elsgolc, L. E., Calculus of Variations, Addison-Wesley (Pergamon Press), Reading, Mass., 1962.
3. Hermsen, R. J., and T. J. Higgins, "Optimal Design of Nuclear Reactor Control Systems by Use of Pontryagin's Maximum Principle," Conference Paper presented at IEEE Salt Lake City meeting, 1964.
4. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Theory," *Trans. ASME, Journal of Basic Engineering*, March, 1969.
5. Kliger, I., "Synthesis of an Optimal Nuclear Reactor Control System," Neutron Dynamics and Control Symposium, CONF 650406, 1966.
6. Kliger, I., "Optimal Control of Space Dependent Nuclear Reactor," *Transactions ANS, Vol. 7, No. 2, June, 1965*.
7. Merriam, C. W., Optimization Theory and the Design of Feedback Control Systems, McGraw-Hill, N. Y., 1964.
8. Mohler, R. R., Optimal Control of Nuclear-Reactor Processes, Ph.D. Dissertation, Michigan, 1965.
9. Mohler, R. R., and J. E. Perry, Jr., "Nuclear Rocket Engine Control," *Nucleonics*, Vol. 19, No. 4, April, 1969.
10. Pontryagin, L. S., et al., The Mathematical Theory of Optimal Processes, Interscience, New York, 1962.
11. Rosztoczy, Z. R., Optimization Studies in Nuclear Engineering, Ph. D. Dissertation, University of Arizona, 1964.
12. Rosztoczy, Z. R., A. P. Sage, and L. E. Weaver, "Application of Pontryagin's Maximum Principle to Flux-State Changes in Nuclear Reactors," TID 7662 AEC Symposium Series 2, Reactor Kinetics and Control, 1964.

13. Secker, P. A., Jr., and S. T. Murray, Computer Code, "Solution of the Time-Varying Matrix Riccati Differential Equation," University of Arizona, August, 1964. (Unpublished)
14. Secker, P. A., Jr., and L. E. Weaver, "Synthesis of Optimal Feedback Control for Nuclear Rocket Engine Start-Up," Neutron Dynamics and Control Symposium, CONF 650406, 1966.
15. Shen, C. N., and F. G. Haag, "Application of Optimal Control to Nuclear Rocket Start-Up," IEEE-TYGNS, Vol. NS-11, No. 2, April, 1964.
16. Weaver, L. E., Systems Analysis of Nuclear Reactor Dynamics, American Nuclear Society and U. S. Atomic Energy Commission Monograph Series on Nuclear Science Technology., Rowman and Littlefield, Inc., New York, 1963.

N66-19884

SECTION II

Modeling With Liapunov Function

## TABLE OF CONTENTS

	Page
LIST OF TABLES . . . . .	v
LIST OF ILLUSTRATIONS . . . . .	vi
ABSTRACT . . . . .	viii
 CHAPTER	
1. INTRODUCTION . . . . .	1
1.1 Introduction and Statement of the Problem . . . . .	1
1.2 Outline of the Work . . . . .	4
2. THE GENERAL MODELING TECHNIQUE . . . . .	6
3. THE SECOND ORDER MODEL . . . . .	17
3.1 The Second Order Model--Method I . . . . .	17
3.2 The Second Order Model--Method II . . . . .	27
3.3 Examples of the Second Order Model . . . . .	31
3.4 Finding Phase Margin by Matching V Curves . . . . .	51
4. THE THIRD ORDER MODEL . . . . .	54
4.1 Development of the Model . . . . .	54
4.2 Examples of the Third Order Model . . . . .	65
5. SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH . . . . .	75
5.1 Summary . . . . .	75
5.2 Suggestions for Further Research . . . . .	76
APPENDIX A: TABULATED <u>P</u> MATRICES FOR GIVEN <u>Q</u> MATRICES . . . . .	82
APPENDIX B: FREQUENCY DOMAIN MODELS . . . . .	86
LIST OF REFERENCES . . . . .	88



LIST OF TABLES

Table	Page
3-1 Results of Example 3-1 . . . . .	40
3-2 Results of Example 3-2 . . . . .	40
3-3 Results of Example 3-3 . . . . .	41
3-4 Results of Example 3-4 . . . . .	41
3-5 Results of Example 3-5 . . . . .	42
4-1 Results of Example 4-1 . . . . .	71
4-2 Results of Example 4-2 . . . . .	72

## LIST OF ILLUSTRATIONS

Figure	Page
2-1 The V Surface (Curve) for a Second Order System . . . . .	9
2-2 V Surfaces of a Third Order System and Its Model . . . . .	12
2-3 Geometrical Illustration of the Matching Technique . . . . .	15
3-1 Rotation of the Coordinate Axes . . . . .	23
Root Locus Plots of the Five Types of Third Order Systems	
3-2 Case 1 . . . . .	33
3-3 Case 2 . . . . .	33
3-4 Case 3 . . . . .	34
3-5 Case 4 . . . . .	34
3-6 Case 5 . . . . .	34
3-7 Transient Responses of the Given System and Model for Example 3-1 . . . . .	43
3-8 Transient Responses of the Given System and Model for Example 3-2 . . . . .	44
3-9 Transient Responses of the Given System and Model for Example 3-3 . . . . .	45
3-10 Transient Responses of the Given System and Model for Example 3-4 . . . . .	46
3-11 Transient Responses of the Given System and Model for Example 3-5 . . . . .	47

LIST OF ILLUSTRATIONS--Continued

	Page
4-1 Transient Responses of the Given System and Models for Example 4-1 . . . . .	73
4-2 Transient Responses of the Given System and Models for Example 4-2 . . . . .	74

ABSTRACT

19884

The sole purpose of this work is to develop a technique by which lower order models can be obtained for linear control systems. This technique is based entirely on Liapunov V functions and thus serves to extend the usefulness of Liapunov functions from strict stability analysis to relative stability analysis. Previously, Liapunov V functions could serve only to determine the stability or instability of a system by considering the nature of the surface described in the state space by this V function.

With this modeling technique, relative stability can also be determined by observing the behavior of the model. Using this technique, models of any order can be developed for a given control system. Of particular interest in the field of analysis are the second and third order models.

For systems with no zeros the second order model and the phase margin model are shown to be approximately equivalent. This is a highly interesting result since the models are based on two entirely different philosophies-- one is developed in the time domain and the other in the frequency domain.

The third order model developed from Liapunov V functions is a better approximation to high-order systems than the phase margin and second order models. It is conceivable that a fourth (or higher) order model would give an even better approximation but the amount of work necessary to obtain this model makes it impractical. Unfortunately, systems with zeros complicate the modeling process. While the method is still valid and workable, difficulty arises in interpreting the results of the method. Chapter 5 discusses this problem.

*Author*

## CHAPTER 1

### INTRODUCTION

#### 1.1 Introduction and Statement of the Problem

In the present state of the art, the most successful and widely used methods of analyzing linear control systems are fabricated on the frequency domain. Such methods are phase margin, M-peak, M-circles, and root locus.<sup>1</sup> Recently however, considerable interest has been shown in finding a method which is entirely dependent on the time domain and in particular the Second Method of Liapunov.<sup>2, 4, 6, 7</sup>

One such method considers upper and lower bounds placed on the Liapunov  $V$  function:<sup>7</sup>

$$V[\underline{x}(0)] e^{-\alpha_1 t} \leq V[\underline{x}(t)] \leq V[\underline{x}(0)] e^{-\alpha_2 t}$$

When  $\alpha = \alpha_1 = \alpha_2$  and  $\alpha$  is a linear combination of the eigenvalues of the system, then equality holds. Unfortunately, analyzing the system with  $V[\underline{x}(t)] = V[\underline{x}(0)] e^{-\alpha t}$  is essentially equivalent to finding the exact solution of the given system and hence this method, while interesting, is of little value. Further, if  $\alpha_1$  and  $\alpha_2$  are not combinations of the eigenvalues, then the bounds may be too loose to give a good indication of the system behavior. Another

approach proposed by Rekasius<sup>6</sup> utilizes a model as an optimum system to measure the performance of higher order systems. However, the use of this method has not been too successful.

Thus all of the existing methods based on the Second Method have been particularly disappointing. This paper presents a new method which, it is hoped, will fill the gap that now exists. This method determines a model which is an approximation to a completely specified system. The model can then easily be analyzed to determine the response of the system.

To find this lower order model, a method of fixing the free parameters of the model must be found. Such a method already exists which used frequency information; it is called the phase margin model (Appendix B). By matching frequency and phase margin at crossover of a second order model to that of a given system, the model is completely specified and can be used to determine the behavior of a given system. Compensation of a control system is accomplished by picking a phase margin model with the desired characteristics (phase margin and crossover frequency, for example), and then adjusting the parameters of a compensation network to give a system with the same phase margin and crossover frequency as the model. In effect, the system which yields the desired model has been found.

Contrary to the phase margin technique which utilizes frequency domain information, the method offered here depends only on time domain information. In fact, the method operates directly on the differential equation describing the given linear system and yields a lower order differential equation as the model. This method is called the V surface modeling technique; and, as its name indicates, it utilizes Liapunov V functions.

The first advantage of the V surface model over the phase margin model might be esthetic, in that the model is entirely dependent on the time domain where the performance of any system must be finally analyzed and not on the artificial frequency domain. Second, the method is readily applicable to digital computers. Third, while the phase margin techniques can yield only a second order model, the V surface modeling technique yields a second order model which is similar to the phase margin model and also any higher order model.

In particular the third order model is investigated in some detail, and is shown to give better results than the phase margin model (and the second order model). Finally, it may be of some interest that Liapunov functions may be used in the analysis and synthesis of linear control systems and not merely in determining their stability.



## 1.2 Outline of the Work

In the following discussion, this V surface modeling technique is developed for various types of systems and then compared to the existing phase margin technique. Chapter 2 is a brief introduction to the concept of the V surface modeling method and gives the general procedure for finding a kth order model of an nth order system ( $k < n$ ).

Chapter 3 deals exclusively with second order models. A second order model is developed in Section 3.1 along the lines suggested in Chapter 2. Unfortunately this model, while perfectly feasible and workable, has a disadvantage in that a matrix must be transformed into its diagonal form. To eliminate this problem Section 3.2 offers another slightly different method for finding second order models. This latter method is used in Section 3.3 to find the second order models for five third order systems. The results of these examples show that the second order model and the phase margin model give almost equivalent models and Section 3.4 attempts to show how the phase margin model and the second order model are related.

Because the second order model gives only fair approximations in some cases, a third order model is found in Chapter 4. This model is developed in Section 4.1 along the lines suggested in Chapter 2 and is used in Section 4.2 to model two fourth order systems. The third order model is seen from these two examples to closely approximate the

actual system and to give much better results than either the second order model or the phase margin model. Throughout Chapters 2 through 4 it is assumed that the systems treated contain no zeros.

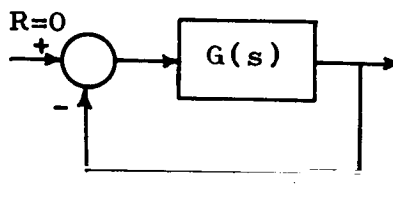
Section 5.2 of Chapter 5 suggests possible schemes to handle systems with zeros as well as suggesting other areas for further research. Section 5.1 is a brief summary of the modeling method and its results.

## CHAPTER 2

### THE GENERAL MODELING TECHNIQUE

In this chapter, an intuitive basis for the modeling technique is first offered and then the general procedure for finding a  $k$ th order model of an  $n$ th order system ( $k < n$ ) is developed.

An  $n$ th order unforced system is completely specified by its characteristic equation if its open loop transfer function  $G(s)$  has the following form:


$$G(s) = \frac{a_1}{s(s^{n-1} + a_n s^{n-2} + \dots + a_3 s + a_2)} \quad (2-1)$$

The characteristic equation of this system is

$$\frac{d^n x}{dt^n} + a_n \frac{d^{n-1} x}{dt^{n-1}} + \dots + a_2 \frac{dx}{dt} + a_1 x = 0$$

Defining  $x = x_1$ ,  $\frac{dx}{dt} = x_2$ ,  $\dots$ ,  $\frac{d^{n-1} x}{dt^{n-1}} = x_n$ , the differential equation can be rewritten as

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \end{aligned}$$

$$\dot{x}_3 = x_4$$

.

.

.

$$\dot{x}_{n-1} = x_n$$

$$\dot{x}_n = -a_1 x_1 - a_2 x_2 - \dots - a_n x_n$$

In vector matrix form the phase variables take the form

$$\underline{\dot{x}} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -a_1 & -a_2 & \dots & \dots & -a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \underline{Ax} \quad (2-2)$$

Thus the unforced system is completely defined by  $\underline{\dot{x}} = \underline{Ax}$  when the system is in the form of Eq. (2-1). When the system has a different form, then the system matrix  $\underline{A}$  does not uniquely determine the system.<sup>2</sup> In this case, a different set of variables other than phase variables must be used to describe the system.

The Liapunov function for an nth order system is defined as

$$V(t) = p_{11}x_1^2 + 2p_{12}x_1x_2 + p_{22}x_2^2 + \dots + p_{nn}x_n^2 = \underline{x}^T \underline{P} \underline{x}$$

where  $\underline{x}$  is the state vector and  $\underline{P}$  is a symmetric matrix.

If the system is linear and autonomous and described by  $\dot{\underline{x}} = \underline{A}\underline{x}$ , then

$$\begin{aligned} \frac{dv}{dt} &= \frac{d}{dt} (\underline{x}^T \underline{P} \underline{x}) = \dot{\underline{x}}^T \underline{P} \underline{x} + \underline{x}^T \underline{P} \dot{\underline{x}} \\ &= \underline{x}^T \underline{A}^T \underline{P} \underline{x} + \underline{x}^T \underline{P} \underline{A} \underline{x} = \underline{x}^T (\underline{A}^T \underline{P} + \underline{P} \underline{A}) \underline{x} = -\underline{x}^T \underline{Q} \underline{x} \end{aligned}$$

and  $\dot{V}(t) = -\underline{x}^T \underline{Q} \underline{x}$  where  $\underline{Q}$  is a symmetric matrix. The elements of  $\underline{P}$  can be found when  $\underline{A}$  and  $\underline{Q}$  are given. At a particular instant of time,  $V(t)$  is equal to some numerical value (say  $V(t) = K$ ) which is determined from the  $x_i$ 's ( $i = 1, 2, \dots, n$ ) by  $V(t_1) = \underline{x}^T \underline{P} \underline{x} = K$ . The value of  $\underline{x}$  which gives  $V(t_1) = K$  is far from unique, in fact, upon plotting  $\underline{x}^T \underline{P} \underline{x} = K$  in the state space ( $x_1, x_2, \dots, x_n$  as coordinates), a surface is obtained, any point of which gives an admissible value of  $\underline{x}$ . Fig. 2-1 shows this surface (curve) when  $n = 2$ . For a positive definite or semi definite  $\underline{Q}$  matrix chosen for  $\dot{V}$ , the surface must obey certain requirements. If the system is unstable then the  $V$  surface is open (hyperbola for  $n = 2$ , etc.), and if the system is stable then the  $V$  surface is closed (ellipse for  $n = 2$ , ellipsoid for  $n = 3$ , etc.). Historically, the stability of a control system was determined by whether the Liapunov  $V$  curve was an open or closed curve (surface for  $n = 3$ , hypersurface for  $n > 3$ ).

But, no indication of relative stability of the system was obtained by regarding the closedness (openness)

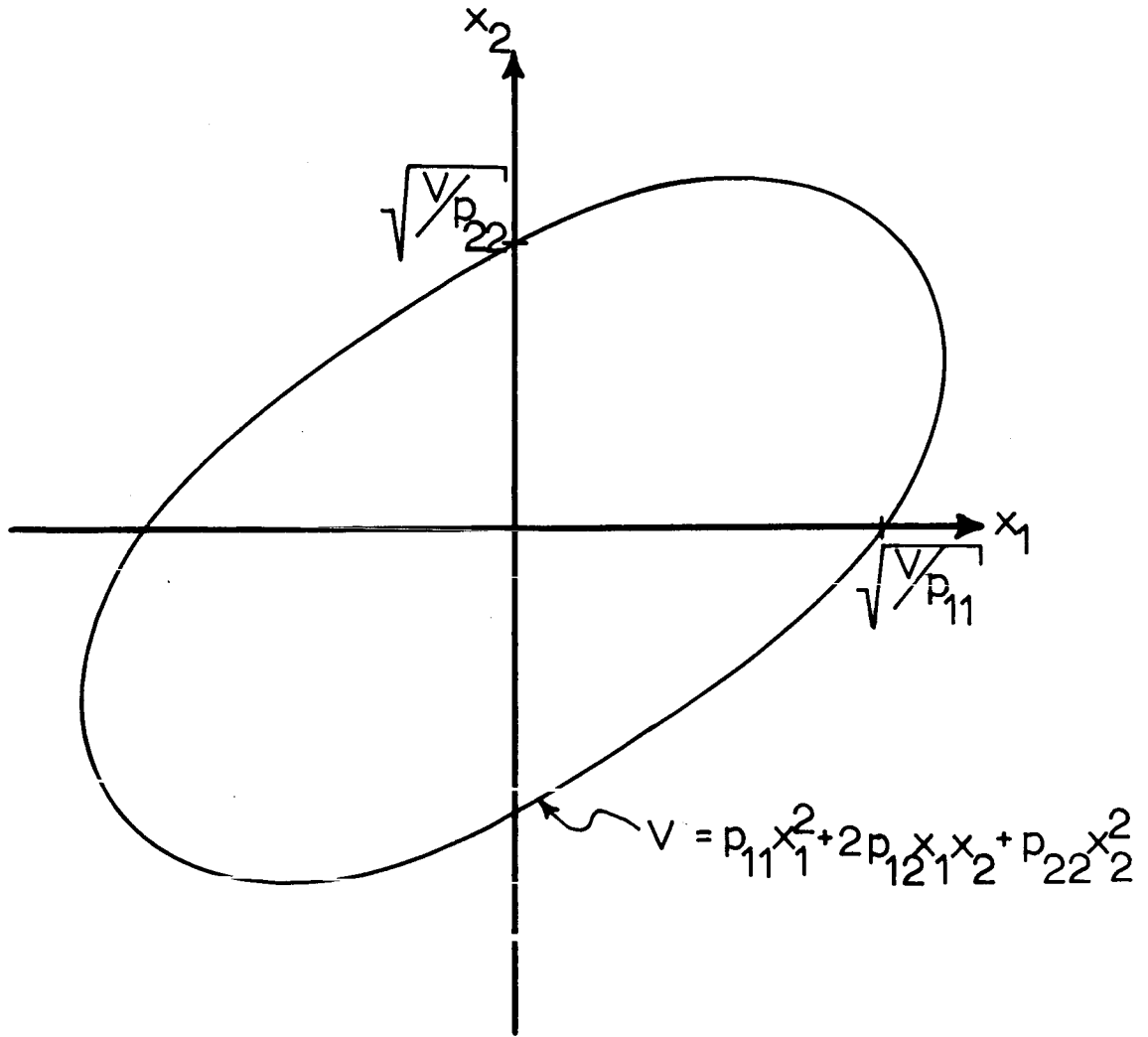


FIG. 2-1

V SURFACE (CURVE) FOR A  
SECOND ORDER SYSTEM

of the  $V$  surface. Such an indication can be obtained, however, by considering the shape of the  $V$  surface. It is intuitive that if two systems have nearly identical responses, then their  $V$  surfaces must have very nearly the same shape and orientation for the same  $\dot{V}$ . Furthermore, if the  $V$  surface of two control systems have exactly the same shape and orientation for the same  $Q$ , then they must be identical systems; since there is term wise equivalence between the two expressions for  $V$ , the  $P$  matrices must be the same and since  $P$  is unique for a given  $A$ , then the two  $A$  matrices must be identical. Thus a measure of the relative stability of a system can be obtained by comparing its  $V$  surface with the  $V$  surface of a system (model) whose relative stability (response to a step input) is known. If the two curves are closely matched, then the two systems can be said to have the same behavior.

It is obvious that the model should be of lower order than the given system, since the best model of the same order as the given system would be the system. A model of higher order would be meaningless. In addition, the model should be of low order (second or third) in order to be easily analyzed.

The objective is to find a model of lower order than the given system. Since a model of  $k$ th order has a  $V$  surface in  $k$  space, and the  $n$ th order model has its  $V$  surface in  $n$  space it is obviously impractical to attempt

to match the  $V$  surfaces since  $k < n$ . However, if two systems have identical responses, then their  $V$  surfaces must be matched in  $k$  space as well as in  $n$  space. Thus to obtain a  $k$ th order model of an  $n$ th order system the procedure is to match the  $V$  surfaces in  $k$  space. See Fig. 2-2. The  $k$ th order system, which gives the matching  $V$  surface in  $k$  space, is called the model, and its response hopefully gives a good approximation of the actual response.

For an  $n$ th order system the  $V$  surface in  $n$  space for a particular  $\dot{V} = -\underline{x}^T \underline{Q} \underline{x}$  is given by

$$V = \underline{x}^T \underline{P} \underline{x} = p_{11}x_1^2 + 2p_{12}x_1x_2 + \dots + p_{kk}x_k^2 + \dots + p_{nn}x_n^2$$

The intersection of this surface in  $k$  space is given by

$$V = p_{11}x_1^2 + 2p_{12}x_1x_2 + \dots + p_{kk}x_k^2 = \underline{x}^k T \underline{P}^k \underline{x}^k$$

where  $\underline{x}^k$  is a vector of the first  $k$  elements of the  $\underline{x}$  vector and  $\underline{P}^k$  is the upper left hand  $k$  by  $k$  submatrix of the  $n$  by  $n$   $\underline{P}$  matrix. It is important not to confuse  $\underline{x}^k$  and  $x_i^2$ .  $x_i^2$  is the square of the  $i$ th component of the vector  $\underline{x}^k$ .



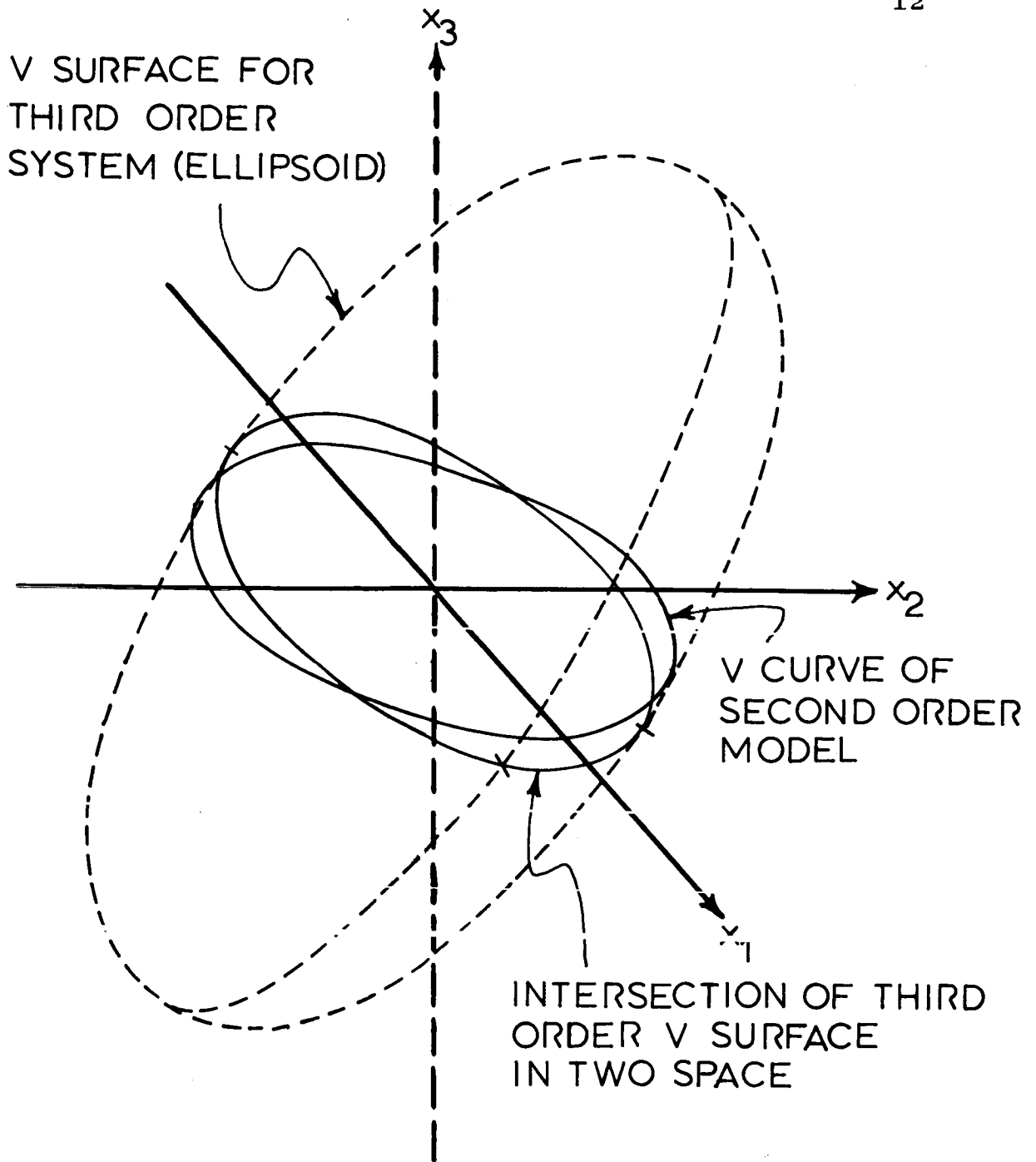


FIG. 2-2

V SURFACES OF A THIRD ORDER  
SYSTEM AND ITS MODEL

$$\underline{P} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1k} & \dots & p_{1n} \\ p_{12} & & & \cdot & & \\ p_{13} & & & \cdot & & \\ \cdot & & & \cdot & & \\ \cdot & & & \cdot & & \\ p_{1k} & \dots & & p_{kk} & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ p_{1n} & & & & & p_{nn} \end{bmatrix}$$

$\underline{p}^k$

For the  $k$ th order model the  $V$  surface for the same  $\dot{V}$  is given by

$$V = \underline{x}^T \underline{M} \underline{x} = m_{11} x_1^2 + 2m_{12} x_1 x_2 + \dots + m_{kk} x_k^2$$

To match the two  $V$  surfaces it is necessary that

$$m_{11} = p_{11}$$

$$m_{12} = p_{12}$$

.

.

.

$$m_{kk} = p_{kk}$$

But these are  $\frac{k(k+1)}{2}$  conditions which must be met by a  $k$ th order system with only  $k$  variables; an impossible feat. It is possible, however, to perfectly match the shapes of the two  $V$  surfaces (they will not have the same orientation) by rotating the axes around until all cross products are eliminated. The diagonal terms of the two transformed matrices may then be set equal to determine the  $k$  parameters of the  $k$ th order model (see Fig. 2-3). One criterion of the "goodness" of the model is how well the orientations of the  $V$  surfaces match. Obviously many different models can be found by using different  $\dot{V}$ 's. If the modeling technique is a good one, then the model and the given system should have much the same response, and thus the  $V$  surfaces should be closely matched for any  $\dot{V}$ . Therefore, the choice of a particular  $\dot{V}$  should not affect the model too much. The ultimate choice of a  $\dot{V}$  should be based on the amount of work that must be expended to find the model.

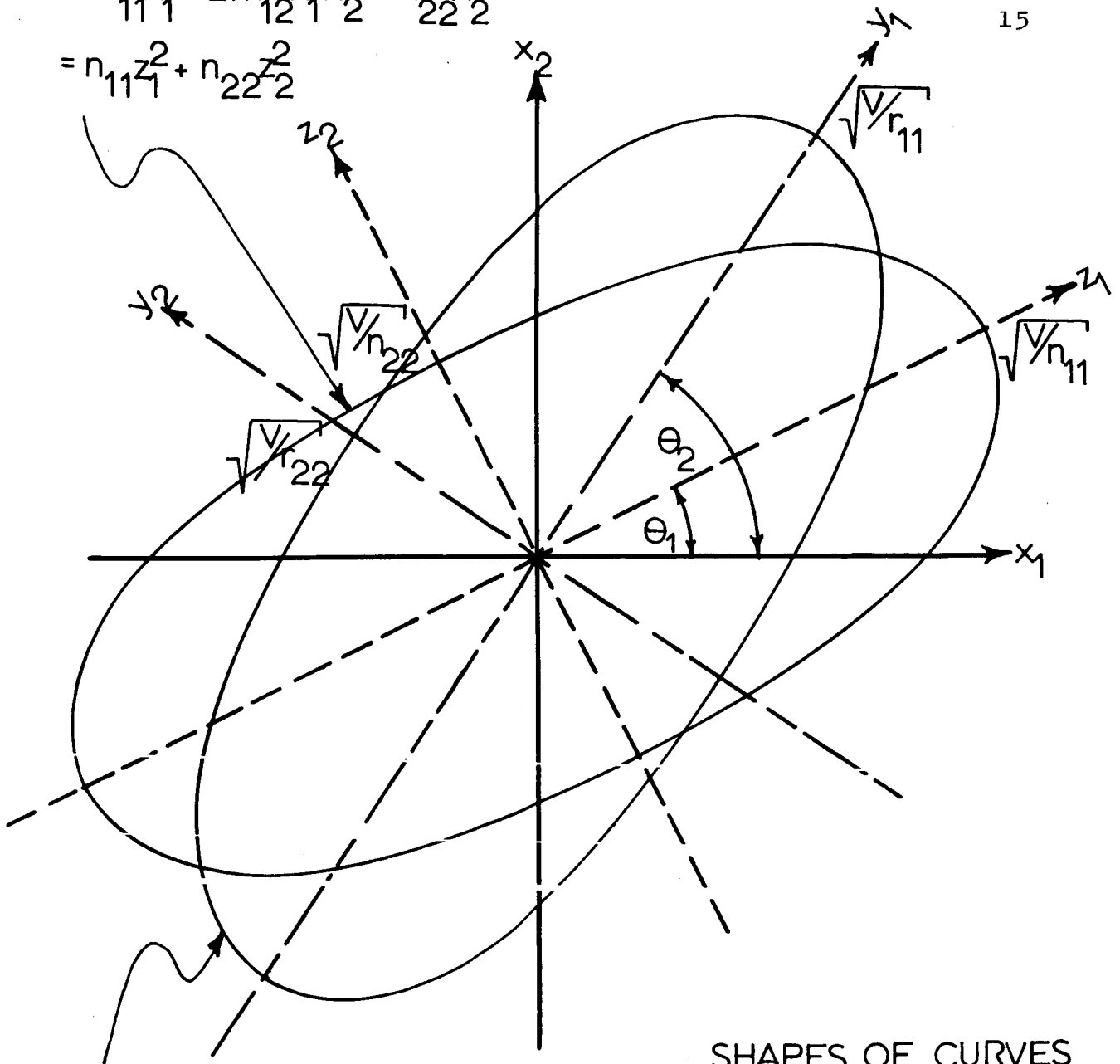
A step by step procedure to find a general  $k$ th order model for an  $n$ th order system is then:

- 1) Pick one  $\dot{V}$  for the two systems.
- 2) Determine  $\underline{M}$  ( $V = \underline{x}^T \underline{Mx}$ ) for the  $k$ th order model and  $\underline{P}$  ( $V = \underline{x}^T \underline{Px}$ ) for the  $n$ th order system for this  $\dot{V}$ .
- 3) Take the upper left hand  $k$  by  $k$  submatrix of  $\underline{P}$  for  $\underline{P}^k$ .
- 4) Eliminate cross terms in  $\underline{P}^k$  and  $\underline{M}$  by a suitable change of variables to rotate the axes.

V CURVE OF MODEL

$$V = m_{11}x_1^2 + 2m_{12}x_1x_2 + m_{22}x_2^2$$

$$= n_{11}z_1^2 + n_{22}z_2^2$$



TWO SPACE INTERSECTION  
OF THIRD ORDER V SURFACE

$$V = p_{11}x_1^2 + 2p_{12}x_1x_2 + p_{22}x_2^2$$

$$= r_{11}y_1^2 + r_{22}y_2^2$$

SHAPES OF CURVES  
MATCHED WHEN

$$n_{11} = r_{11}$$

$$n_{22} = r_{22}$$

FIG. 2-3

GEOMETRICAL ILLUSTRATION OF  
THE MATCHING TECHNIQUE

- 5) Adjust the parameters of the  $k$ th order system to match the diagonal elements of the transformed matrices.

In the following chapters the second order model ( $k = 2$ ) and the third order model ( $k = 3$ ) are investigated in detail. Some examples are worked to show the application of the technique, and the results are compared to those obtained by phase margin techniques.

## CHAPTER 3

### THE SECOND ORDER MODEL

The generalized procedure for finding a kth order model offered in Chapter 2 is now applied to the second order case ( $k = 2$ ). As seen in Section 3.1, a certain disadvantage (rotation of axes) results using this procedure and thus an alternate method is given in Section 3.2. Several examples are given in Section 3.3 to illustrate the method and to show how well it approximates the actual system. This second order model is also compared with the phase margin model and it is very satisfying to note that the two are approximately the same for all cases investigated.

#### 3.1 The Second Order Model--Method I

To obtain a suitable second order model for a system, first a  $\dot{V}$  must be chosen for both the model and the given system. Since the  $\dot{V}$  should be applicable to both the model and the system, it should depend only on variables which are defined for both systems ( $x_1$  and  $x_2$  for a second order model). Thus in general  $\dot{V} = q_{11}x_1^2 + 2q_{12}x_1x_2 + q_{22}x_2^2$ . The V curve (ellipse) for the second order model then becomes  $V = \underline{x}^T \underline{M} \underline{x}$  where

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \underline{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{bmatrix}$$

$\underline{M}$  is obtained from the matrix equation  $\underline{A}^T \underline{M} + \underline{M} \underline{A} = -\underline{Q}$ , where  $\underline{A}$  is the system matrix of the second order model, and

$$\underline{Q} = \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix}$$

The V curve for the given nth order system is  $V' = \underline{x}^T \underline{P} \underline{x}$  where

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \hat{x}_n \end{bmatrix}, \quad \underline{P} = \begin{bmatrix} p_{11} & p_{12} & \cdot & \cdot & \cdot & p_{1n} \\ p_{12} & p_{22} & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \hat{p}_{1n} & \cdot & \cdot & \cdot & & \hat{p}_{nn} \end{bmatrix}$$

$\underline{P}$  is obtained from the matrix equation  $\underline{A}'^T \underline{P} + \underline{P} \underline{A}' = -\underline{Q}'$ , where  $\underline{A}'$  is the system matrix of the given system, and

$$\underline{Q}' = \begin{bmatrix} q_{11} & q_{12} & 0 & \cdot & \cdot & \cdot & 0 \\ q_{12} & q_{22} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & & & & & \cdot \\ \cdot & \cdot & & & & & \cdot \\ \cdot & \cdot & & & & & \cdot \\ \cdot & \cdot & & & & & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}$$

Notice that the double use of  $x_1$  and  $x_2$  for both the system and the model specifies that the first two variables of the system and the two variables of the model must be the same; i.e., they must have identical meaning for the system and the model. In this instance, for example,  $x_1$  and  $x_2$  are the output of the system (or model) and  $\dot{x}_2$  its first derivative.

The 2 space intersection (intersection in  $x_1, x_2$  plane--an ellipse) of the curve in  $n$  space ( $V' = \underline{x}^T \underline{P} \underline{x}$ ) is

$$V' = \underline{x}^{2T} \underline{P}^2 \underline{x}^2 \text{ where } \underline{x}^2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \underline{P}^2 = \begin{bmatrix} P_{11} & P_{12} \\ P_{12} & P_{22} \end{bmatrix}$$

To match the curves  $V'$  and  $V$ , set the expressions for these two curves equal and equate like terms.

$$m_{11} = p_{11}$$

$$m_{12} = p_{12}$$

$$m_{22} = p_{22}$$

The elements of  $\underline{P}^2$  are fixed by the given system and  $\dot{V}$ , and the two parameters of the model must be adjusted so that the elements of  $\underline{M}$  match the elements of  $\underline{P}^2$ . But it is impossible to satisfy three equations with only two variable parameters. Therefore, the curves  $V'$  and  $V$  cannot be exactly matched. It is possible however to match the



shapes of the two curves while ignoring the orientations of the two ellipses. This is accomplished by matching the major and minor axes of the ellipses.

Under a suitable change of variables, it is possible to rotate the coordinate axes and eliminate cross terms. For  $V = \underline{x}^T \underline{M} \underline{x}$  then, it is possible to find some transformation  $\underline{x} = \underline{B} \underline{z}$  such that  $V = \underline{z}^T \underline{B}^T \underline{M} \underline{B} \underline{z}$ , where  $\underline{N} = \underline{B}^T \underline{M} \underline{B}$  is a diagonal matrix.

$$\underline{N} = \begin{bmatrix} n_{11} & 0 \\ 0 & n_{22} \end{bmatrix}$$

The major and minor axes are given by  $2\sqrt{V/n_{22}}$  and  $2\sqrt{V/n_{11}}$  (see Fig. 2-3). Similarly for  $V' = \underline{x}^{2T} \underline{P}^2 \underline{x}^2$  there exists a transformation  $\underline{x}^2 = \underline{C} \underline{y}$  such that  $V' = \underline{y}^T \underline{C}^T \underline{P}^2 \underline{C} \underline{y} = \underline{y}^T \underline{R} \underline{y}$  where  $\underline{R} = \underline{C}^T \underline{P}^2 \underline{C}$  is a diagonal matrix.

$$\underline{R} = \begin{bmatrix} r_{11} & 0 \\ 0 & r_{22} \end{bmatrix}$$

To insure that the final ellipses have the same size, set  $V' = V$  (this does not affect the shape of the ellipses). Then the major and minor axes are given by  $2\sqrt{V/r_{11}}$  and  $2\sqrt{V/r_{22}}$ . To match the shapes of the two curves, set the major and minor axes of the two ellipses equal or,

$$2\sqrt{V/n_{11}} = 2\sqrt{V/r_{22}}$$

$$2\sqrt{V/n_{22}} = 2\sqrt{V/r_{11}}$$

The necessary condition to match the axes is that  $n_{11} = r_{11}$  and  $n_{22} = r_{22}$  or  $\underline{N} = \underline{R}$ . It is now imperative to determine exactly how this rotation of axes is to be accomplished.

For the quadratic form  $\underline{u}^T \underline{S} \underline{u}$ ,

$$\underline{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad \underline{S} = \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix}$$

find the change of coordinates  $\underline{u} = \underline{B} \underline{v}$ ,

$$\underline{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

such that  $\underline{u}^T \underline{S} \underline{u} = \underline{v}^T \underline{B}^T \underline{S} \underline{B} \underline{v} = \underline{v}^T \underline{T} \underline{v}$  where  $\underline{T} = \underline{B}^T \underline{S} \underline{B}$  is a diagonal matrix.

$$\underline{T} = \begin{bmatrix} t_{11} & 0 \\ 0 & t_{22} \end{bmatrix}$$

The requirement on this coordinate transformation is that it be orthogonal (distance preserving); therefore the transformation change must be a rotation of the coordinate

axes. From Fig. 3-1 it is seen that B must be of the form

$$\underline{B} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

and the next step is to find  $\theta$  and the elements of  $\underline{T}$ .

$$\begin{aligned} \underline{T} = \underline{B}^T \underline{S} \underline{B} &= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} (s_{11} \cos^2\theta + 2s_{12} \cos\theta \sin\theta + s_{22} \sin^2\theta) \\ (\cos\theta \sin\theta (s_{22} - s_{11}) + s_{12} (\cos^2\theta - \sin^2\theta)) \\ (\cos\theta \sin\theta (s_{22} - s_{11}) + s_{12} (\cos^2\theta - \sin^2\theta)) \\ (s_{11} \sin^2\theta - 2s_{12} \cos\theta \sin\theta + s_{22} \cos^2\theta) \end{bmatrix} \end{aligned}$$

Therefore,

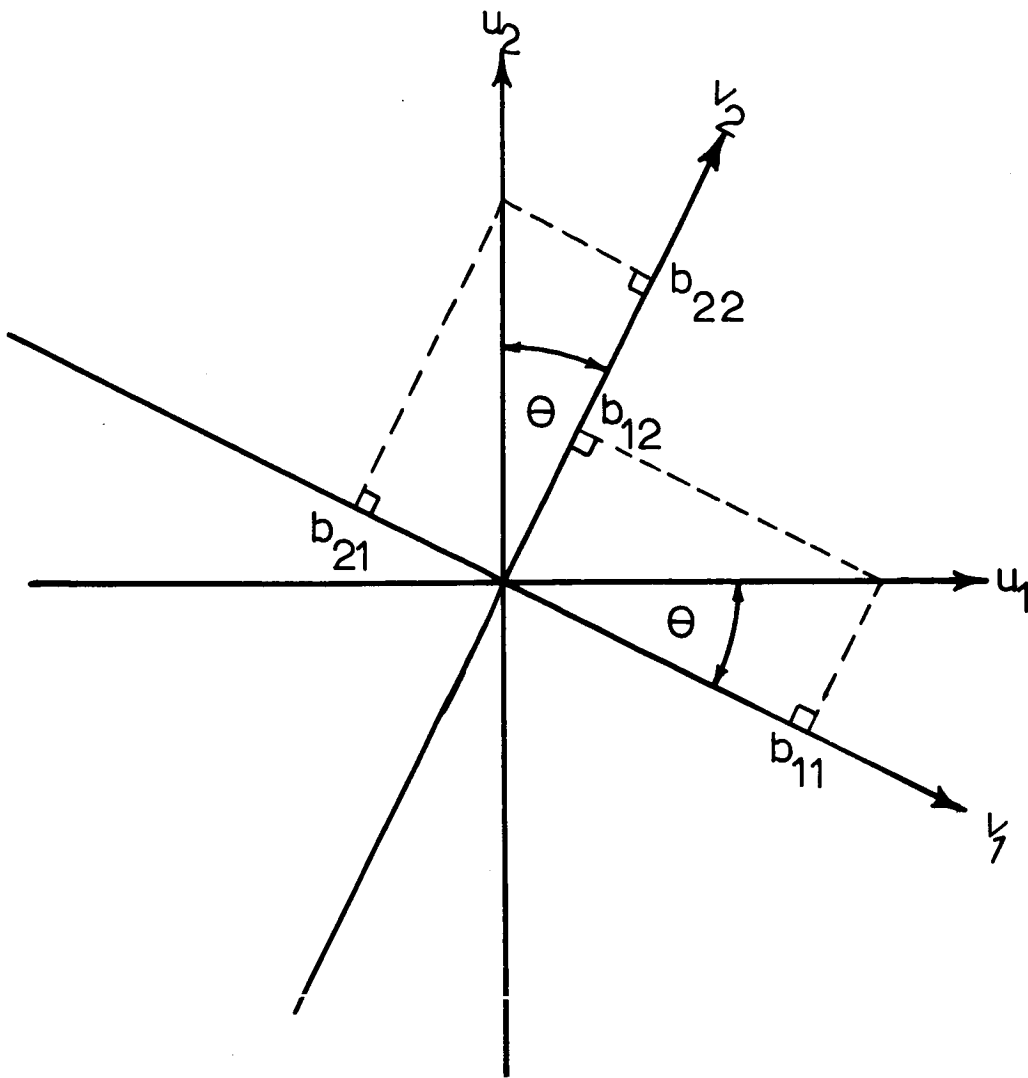
$$\begin{aligned} t_{11} &= s_{11} \cos^2\theta + 2s_{12} \cos\theta \sin\theta + s_{22} \sin^2\theta \\ t_{22} &= s_{11} \sin^2\theta - 2s_{12} \cos\theta \sin\theta + s_{22} \cos^2\theta \end{aligned} \quad (3-1)$$

also,

$$t_{12} = 0 = \cos\theta \sin\theta (s_{22} - s_{11}) + s_{12} (\cos^2\theta - \sin^2\theta)$$

$$\left(\frac{s_{11} - s_{22}}{2}\right) 2\cos\theta \sin\theta = s_{12} (\cos^2\theta - \sin^2\theta)$$

$$\left(\frac{s_{11} - s_{22}}{2}\right) \sin 2\theta = s_{12} \cos 2\theta$$



$$\underline{u} = \underline{\bar{B}}\underline{v}$$

$$u_1 = b_{11}v_1 + b_{12}v_2 = \cos\theta v_1 + \sin\theta v_2$$

$$u_2 = b_{21}v_1 + b_{22}v_2 = -\sin\theta v_1 + \cos\theta v_2$$

$$\underline{\bar{B}} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}, \quad \underline{\bar{B}}^{-1} = \underline{\bar{B}}^T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

FIG. 3-1

ROTATION OF THE COORDINATE AXES

$$\frac{\sin 2\theta}{\cos 2\theta} = \tan 2\theta = \frac{2s_{12}}{s_{11} - s_{22}}$$

$$\text{or } \theta = 1/2 \tan^{-1} \left( \frac{2s_{12}}{s_{11} - s_{22}} \right) \quad (3-2)$$

With these formulas it is possible to transform  $\underline{P}^2$  into its diagonal form  $\underline{R}$  since all elements of  $\underline{P}^2$  are constants and the angle of rotation  $\theta$  can be obtained immediately.

However, the elements of  $\underline{M}$  are not constants, but functions of the variable parameters of the model, and the angle of rotation is not readily available. For example, consider  $\underline{M}$  for  $q_{11} = 1$ .

$$\underline{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$\underline{M}$  is obtained from the matrix equation  $\underline{A}^T \underline{M} + \underline{M} \underline{A} = -\underline{Q}$  where  $\underline{A}$  is the system matrix of the model.

$$\underline{A} = \begin{bmatrix} 0 & 1 \\ -a_1 & -a_2 \end{bmatrix}$$

From part 1 of Appendix A,  $\underline{M}$  turns out to be

$$\underline{M} = \frac{1}{2a_1 a_2} \begin{bmatrix} (a_1 + a_2^2) & a_2 \\ a_2 & 1 \end{bmatrix}$$

To transform  $\underline{M}$  into its diagonal form  $\underline{N}$ , Eqs. (3-1) and (3-2) are used. The elements of  $\underline{N}$  then become:

$$n_{11} = \frac{(a_1 + a_2^2)}{2a_1 a_2} \cos^2 \theta + \frac{2a_2}{2a_1 a_2} \cos \theta \sin \theta + \frac{1}{2a_1 a_2} \sin^2 \theta$$

$$n_{12} = 0$$

$$n_{22} = \frac{(a_1 + a_2^2)}{2a_1 a_2} \sin^2 \theta - \frac{2a_2}{2a_1 a_2} \cos \theta \sin \theta + \frac{1}{2a_1 a_2} \cos^2 \theta$$

where

$$\theta = 1/2 \tan^{-1} \left( \frac{2a_2}{a_1 + a_2^2 - 1} \right)$$

Thus  $\theta$  is given in terms of the parameters of the model and since these parameters are unknown,  $\theta$  is not known.

Therefore, the procedure of matching  $n_{11} = r_{11}$  and  $n_{22} = r_{22}$  must be one of trial and error; unless, of course,  $\underline{M}$  is already in diagonal form and no transformation is needed. Such is the case when  $\dot{V} = -x_2^2$  ( $q_{11} = q_{12} = 0$ ,  $q_{22} = 1$ ). For this  $\dot{V}$ ,  $\underline{M}$  becomes

$$\underline{M} = \begin{bmatrix} \frac{a_1}{2a_2} & 0 \\ 0 & \frac{1}{2a_2} \end{bmatrix}$$

where  $a_1$  and  $a_2$  are the variable parameters of the system matrix  $\underline{A}$ .

$$\underline{A} = \begin{bmatrix} 0 & 1 \\ -a_1 & -a_2 \end{bmatrix}$$

Therefore,  $a_1/2a_2 = r_{11}$  and  $1/2a_2 = r_{22}$

or  $a_2 = 1/2r_{22}$

$$a_1 = r_{11}/r_{22} \quad (3-3)$$

The procedure for finding the second order model for an nth order system is summarized in the following steps:

1. Pick  $\dot{V} = -x_2^2 = -\underline{x}^T \underline{Q} \underline{x}$  ( $q_{22} = 1$ ) and solve for  $\underline{P}$ , an n by n matrix, from the matrix equation  $\underline{A}'^T \underline{P} + \underline{P} \underline{A}' = -\underline{Q}$  where  $\underline{A}'$  is the nth order system matrix.
2. Take the upper left hand 2 by 2 submatrix of  $\underline{P}$  for  $\underline{p}^2$ .
3. Rotate the coordinate axes of the state space using Eqs. (3-1) and (3-2) to find the diagonal form  $\underline{R}$ .

$$\underline{R} = \begin{bmatrix} r_{11} & 0 \\ 0 & r_{22} \end{bmatrix}$$

4. Adjust  $a_1$  and  $a_2$  (parameters of the second order model) to satisfy relation (3-3). The model is now completely specified by  $\underline{A}$ .

$$\underline{A} = \begin{bmatrix} 0 & 1 \\ -a_1 & -a_2 \end{bmatrix}$$

A convenient measure of the quality of the model, i.e., how well it approximates the given system, is given by the angle of rotation  $\theta$ . This angle shows how closely the orientations of the two ellipses match. When  $\theta$  is small ( $\theta < 10^\circ$ , for example) the model is a good approximation to the given system. As  $\theta$  increases the model becomes progressively worse. The disadvantage of this method is that the  $\underline{P}^2$  matrix must be transformed into its diagonal form. While this operation can be readily performed it is desirable to find a method which is more applicable to digital programming.

### 3.2 The Second Order Model--Method II

If the model obtained in the previous manner is a good one, then for any  $\dot{V}$  (not just  $\dot{V} = -x_2^2$ ) the V curve of this model should closely match the V curve intersection of the nth order system. It is conceivable that a slightly better model could be obtained by matching the curves for a different  $\dot{V}$ . Another way to find a model would be a method which attempts to match the curves for more than one  $\dot{V}$ ; i.e., using some criteria, this method would attempt to match V curves for two different choices of  $\dot{V}$  (say  $\dot{V}_1 = -x_1^2$  and  $\dot{V}_2 = -x_2^2$ ). Such a method would attempt to give a



compromise between the model obtained for  $\dot{V}_1 = -x_1^2$  and the model for  $\dot{V}_2 = -x_2^2$  (or any other choices for  $\dot{V}$ ). Hopefully such a model would give a better approximation than the other two. One such method uses, as its criteria, the matching of the products of the major and minor axes of the two ellipses obtained for two different choices of  $\dot{V}$ . From Fig. 2-3 notice that the major and minor axes of the V curve are given by  $2\sqrt{V/r_{11}}$  and  $2\sqrt{V/r_{22}}$  (or  $2\sqrt{V/n_{11}}$  and  $2\sqrt{V/n_{22}}$  for the V curve of the model). The product of these two axes is  $\text{Prod.} = 4V/(r_{11}r_{22})^{1/2}$ . But,  $r_{11}r_{22}$  is the value of the determinant  $|\underline{R}|$  and therefore,  $\text{Prod.} = 4V/|\underline{R}|^{1/2}$ . This  $\underline{R}$  was obtained by an orthogonal transformation of the coordinate axes  $\underline{x} = \underline{C}\underline{y}$ ; giving V as

$$V = \underline{x}^T \underline{P}^2 \underline{x} = \underline{y}^T \underline{C}^T \underline{P}^2 \underline{C} \underline{y} = \underline{y}^T \underline{R} \underline{y}$$

$$\text{where } \underline{R} = \underline{C}^T \underline{P}^2 \underline{C}$$

$$\text{and } |\underline{R}| = |\underline{C}^T \underline{P}^2 \underline{C}| = |\underline{C}^T| |\underline{P}^2| |\underline{C}| = |\underline{C}^T| |\underline{C}| |\underline{P}^2| = |\underline{C}^T \underline{C}| |\underline{P}^2|$$

But since the transformation is orthogonal, then  $\underline{C}^{-1} = \underline{C}^T$  or  $\underline{C}^T \underline{C} = \underline{I}$ . Therefore,  $|\underline{R}| = |\underline{I}| |\underline{P}^2| = |\underline{P}^2|$  and  $\text{Prod.} = 4V/|\underline{P}^2|^{1/2}$ . Similarly for the model the product of the major and minor axes is  $\text{Prod.} = 4V/|\underline{M}|^{1/2}$ .

$$\text{Set } |\underline{M}| = |\underline{P}^2| (m_{11}m_{22} - m_{12}^2 = p_{11}p_{22} - p_{12}^2) \quad (3-4)$$

The axes product is now matched for a particular  $\dot{V}$ . Since the model has two variable parameters, it is possible to

satisfy two such conditions as Eq. (3-4) for two different choices of  $\dot{V}$ . Two convenient  $\dot{V}$ 's are  $\dot{V}_1 = -x_1^2$  ( $q_{11} = 1$ ) and  $\dot{V}_2 = -x_2^2$  ( $q_{22} = 1$ ). For  $\dot{V}_1 = -x_1^2$   $\underline{P}_1$  is obtained from  $\underline{A}'^T \underline{P}_1 + \underline{P}_1 \underline{A}' = -\underline{Q}_1$ , where  $\underline{A}'$  is the nth order system matrix, and  $\underline{Q}_1$  has  $q_{11} = 1$ . The second order model has

$$\underline{M}_1 = \begin{bmatrix} \frac{a_1 + a_2^2}{2a_1 a_2} & \frac{1}{2a_1} \\ \frac{1}{2a_1} & \frac{1}{2a_1 a_2} \end{bmatrix}$$

$$\text{and } |\underline{M}_1| = 1/(4a_1 a_2^2)$$

where  $a_1$  and  $a_2$  are the variable parameters of the system matrix A. Setting  $|\underline{M}_1| = |\underline{P}_1^2|$  ( $\underline{P}_1^2$  is the upper left hand 2 by 2 submatrix of  $\underline{P}_1$ )

$$1/4a_1 a_2^2 = |\underline{P}_1^2| \quad (3-5)$$

For  $\dot{V}_2 = -x_2^2$  ( $q_{22} = 1$ )  $\underline{P}_2$  is obtained from  $\underline{A}'^T \underline{P}_2 + \underline{P}_2 \underline{A}' = -\underline{Q}_2$  where  $\underline{A}'$  is the nth order system matrix and  $\underline{Q}_2$  has  $q_{22} = 1$ . The second order model has

$$\underline{M}_2 = \begin{bmatrix} \frac{a_1}{2a_2} & 0 \\ 0 & \frac{1}{2a_2} \end{bmatrix}$$

$$\text{and } |\underline{M}_2| = a_1/(4a_2^2)$$

Setting  $|\underline{M}_2| = |\underline{P}_2^2|$  ( $\underline{P}_2^2$  is the upper left hand 2 by 2 submatrix of  $\underline{P}_2$ )

$$a_1/4a_2^2 = |\underline{P}_2^2| \quad (3-6)$$

Eqs. (3-5) and (3-6) can be solved simultaneously for  $a_1$  and  $a_2$

$$\begin{aligned} a_1 &= (|\underline{P}_2^2|/|\underline{P}_1^2|)^{1/2} \\ a_2 &= 1/(16|\underline{P}_1^2||\underline{P}_2^2|)^{1/4} \end{aligned} \quad (3-7)$$

The major and minor axes products are now matched for  $\dot{V}_1 = -x_1^2$  and  $\dot{V}_2 = -x_2^2$ . The step by step procedure to find the second order model using this method is as follows:

1. Pick  $\dot{V}_1 = -x_1^2$  and find  $\underline{P}_1$  ( $V_1 = \underline{x}^T \underline{P}_1 \underline{x}$ )
2. Pick  $\dot{V}_2 = -x_2^2$  and find  $\underline{P}_2$  ( $V_2 = \underline{x}^T \underline{P}_2 \underline{x}$ )
3. Take the upper left hand 2 by 2 submatrix of  $\underline{P}_1$  and  $\underline{P}_2$  for  $\underline{P}_1^2$  and  $\underline{P}_2^2$  respectively.
4. Find  $|\underline{P}_1^2|$  and  $|\underline{P}_2^2|$  and use Eqs. (3-7) to find  $a_1$  and  $a_2$ . The model is now completely specified by  $\underline{A}$ .

$$\underline{A} = \begin{bmatrix} 0 & 1 \\ -a_1 & -a_2 \end{bmatrix}$$

This model may or may not be better than the one obtained by matching the curves for one  $\dot{V}$ . The one

advantage it has over the other method is that no rotation of the axes is required; it is simply necessary to find  $\underline{P}_1^2$  and  $\underline{P}_2^2$ . For this reason the latter method is more readily programmed on a digital computer (it is easy to find a single subroutine to give  $|\underline{P}_1^2|$  and  $|\underline{P}_2^2|$ ). A disadvantage of this method is that there exists no basis for determining how good the model is except by actual comparison. In the previous method the angle of rotation  $\theta$  gave this information. Because of its ease of application the last method is used in the examples to follow.

### 3.3 Examples of the Second Order Model

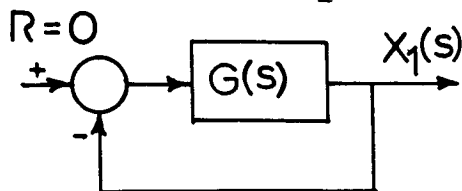
In this section, five third order systems are modeled with the second order model and the phase margin model. The first example is worked out in detail to illustrate the procedure; the results of the other four examples are given to show how the models compare with the given systems. For systems of order greater than three, the procedure is exactly the same, and in a later section two fourth order systems are modeled with second order models. To aid in the hand calculations, an appendix is added at the end of this paper giving tables for finding the  $\underline{P}$  matrices for systems up to fourth order.<sup>3</sup> To model higher order systems a computer is recommended. The five examples in this section were chosen as representatives of the five different types of third order systems of interest:

1. Two complex conjugate poles and one real negative pole far out on the axis (see root locus plot, Fig. 3-2).
2. Two complex conjugate poles and one real negative pole closer to the imaginary axis (Fig. 3-3).
3. Two complex conjugate poles and one real negative pole close to the imaginary axis (Fig. 3-4).
4. One real negative pole close to the imaginary axis and two complex conjugate poles farther from the imaginary axis (Fig. 3-5).
5. Three real negative closed loop poles (Fig. 3-6).

The results of all five examples are given in Tables 3-1 to 3-5, and the transient responses of the given system and the second order model to an initial condition of  $x_1(0) = 1$  are shown in Figs. 3-7 to 3-11.

Example 3-1 (Finding the model for a third order system)

Given the third order system  $G(s)$ , find the second order model and the phase margin model, and compare the transient responses of each to the transient response of the system for  $x_1(0) = 1$ .



$$G(s) = \frac{k}{s(s+a)(s+b)} = \frac{10}{s(s+1)(s+10)}$$

First find the differential equation describing the system from the block diagram.

# ROOT LOCUS PLOTS OF THE FIVE TYPES OF THIRD ORDER SYSTEMS

x = open loop poles of the system

⊠ = closed loop poles of the system

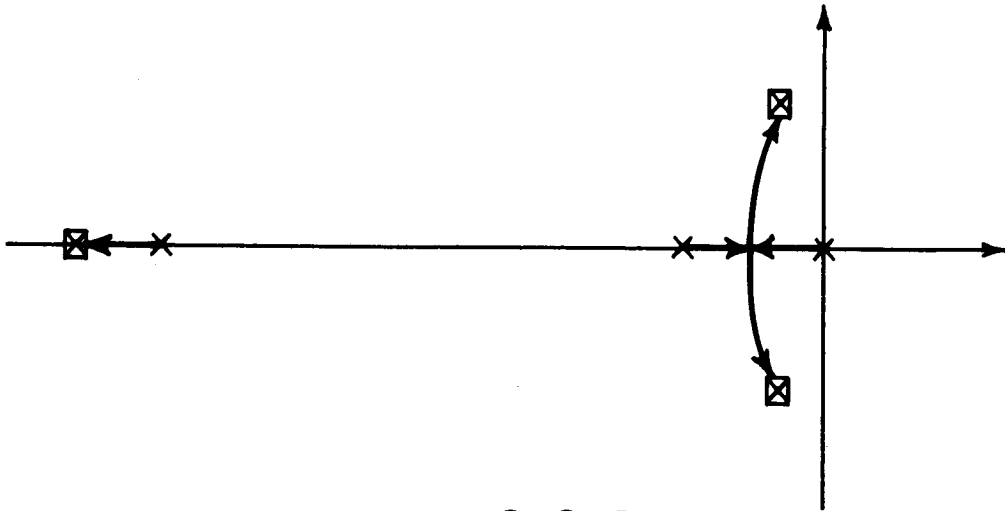


FIG. 3-2  
CASE 1

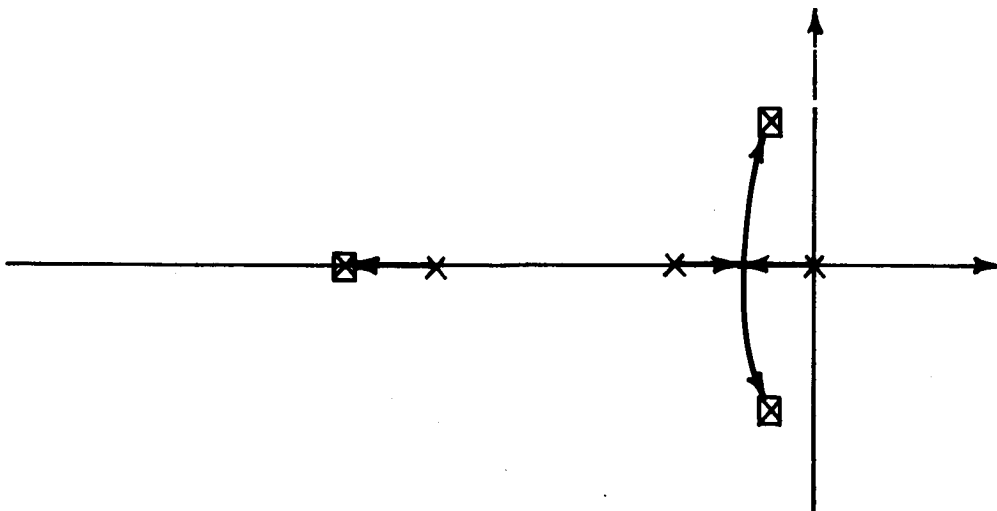


FIG. 3-3  
CASE 2

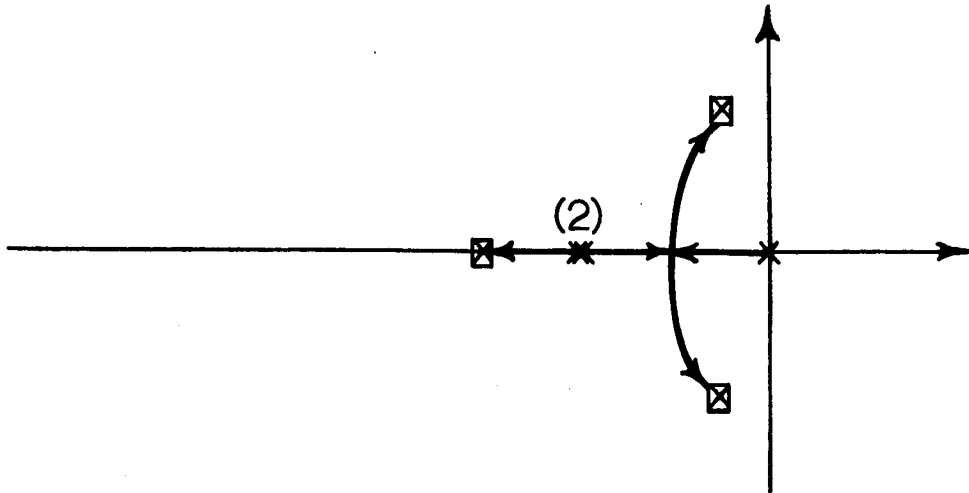


FIG. 3-4  
CASE 3

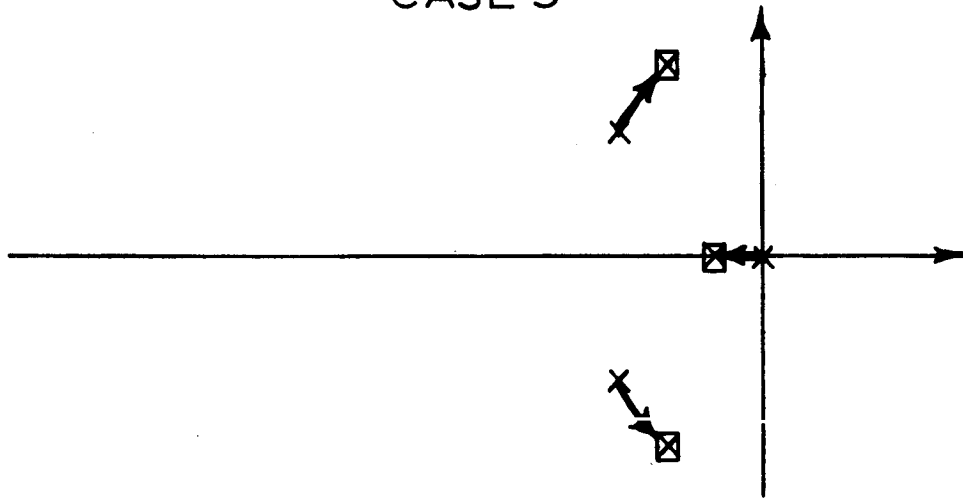


FIG. 3-5  
CASE 4

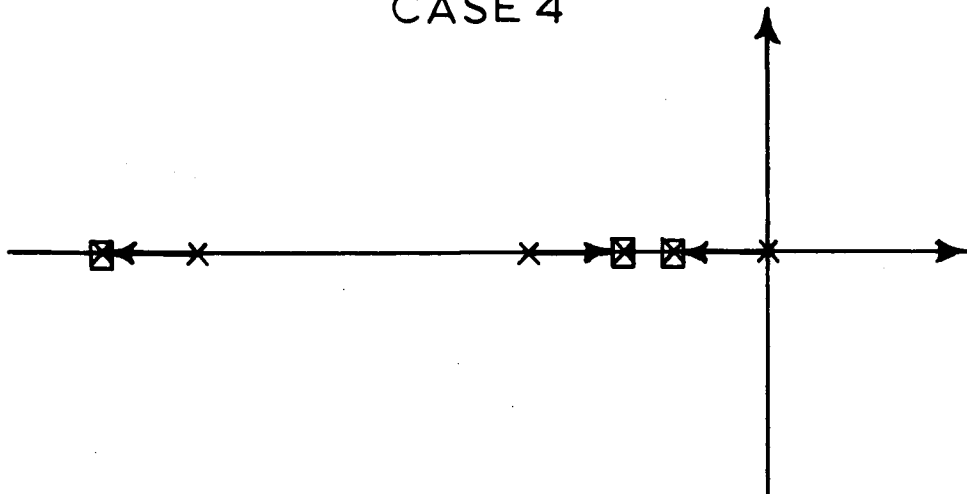


FIG. 3-6  
CASE 5

$$\frac{X_1(s)}{-X_1(s)} = G(s) = \frac{10}{s^3 + 11s^2 + 10s}$$

$$\ddot{x} + 11\dot{x} + 10x = -10x$$

$$x_1 = x$$

defining,  $x_2 = \dot{x}_1 = \dot{x}$

$$x_3 = \dot{x}_2 = \ddot{x}$$

the differential equation can be written as:

$$\dot{x}_3 = -10x_1 - 10x_2 - 11x_3$$

$$\dot{x}_2 = x_3$$

$$\dot{x}_1 = x_2$$

$$\text{or } \underline{\dot{x}} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -10 & -10 & -11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underline{A}' \underline{x}$$

The system transient response,  $x_1(t)$ , for  $x_1(0) = 1$   
 $x_2(0) = x_3(0) = 0$  is:

$$x_1(t) = 1/92 e^{-10.11t} + 91/92 e^{-.445t} \cos.89t \\ + .618 e^{-.445t} \sin.89t$$

Now find the second order model and its transient response for  $x_1(0) = 1$ . (Refer to the procedure at the end of Section 3.2.)



1. Pick  $\dot{V} = -x_1^2 = -\underline{x}^T \underline{Q} \underline{x}$ , then for the third order system

$$\underline{Q}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Using part 2 of the appendix and the system matrix  $\underline{A}'$  for  $\underline{A}$  ( $a_1 = 10$ ,  $a_2 = 10$ ,  $a_3 = 11$ ) then,

$$\underline{P}_1 = \frac{1}{2000} \begin{bmatrix} 2210 & 1210 & 100 \\ 1210 & 1340 & 121 \\ 100 & 121 & 11 \end{bmatrix}$$

2. Pick  $\dot{V} = -x_2^2 = -\underline{x}^T \underline{Q} \underline{x}$ , then

$$\underline{Q}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Using part 2 of Appendix A,

$$\underline{P}_2 = \frac{1}{200} \begin{bmatrix} 110 & 10 & 0 \\ 10 & 131 & 11 \\ 0 & 11 & 1 \end{bmatrix}$$

3. Take the upper left hand 2 by 2 submatrix for  $\underline{P}^2$ , then,

$$\underline{P}_1^2 = \frac{1}{200} \begin{bmatrix} 221 & 121 \\ 121 & 134 \end{bmatrix}, \quad \underline{P}_2^2 = \frac{1}{200} \begin{bmatrix} 110 & 10 \\ 10 & 131 \end{bmatrix}$$

$$4. \quad |P_{-1}^2| = (1/200)^2((221)(134) - (121)^2) = 1.5/4$$

$$|P_{-2}^2| = (1/200)^2((110)(131) - (10)^2) = 1.43/4$$

Using Eqs. (3-7) to find the parameters of the system.

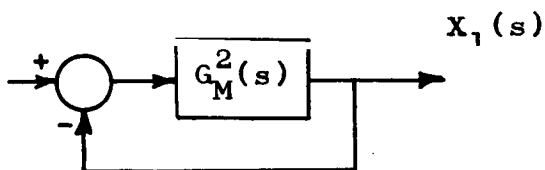
$$a_1 = (|P_{-2}^2| / |P_{-1}^2|)^{1/2} = .972$$

$$a_2 = 1/(16 |P_{-1}^2| |P_{-2}^2|)^{1/4} = .825$$

The system matrix for the model is thus

$$\underline{A} = \begin{bmatrix} 0 & 1 \\ -.972 & -.825 \end{bmatrix}$$

It is convenient to choose the model in the form of  $G_M^2(s) = \frac{k}{s(s+a)}$  where  $G_M^2(s)$  is the open loop transfer function of the second order model.



For  $G_M^2(s) = \frac{k}{s(s+a)}$  the system matrix is

$$\underline{A} = \begin{bmatrix} 0 & 1 \\ -k & -a \end{bmatrix}$$

Therefore,  $k = a_1$ ,  $a = a_2$  and

$$G_M^2(s) = \frac{.972}{s(s+.825)}$$

For a second order system of the form  $\frac{k}{s(s+a)}$ , the transient response for an initial condition  $x_1(0) = 1$  can immediately be written as:

$$x_1(t) = e^{-a/2t} \cos \sqrt{k-a^2/4}t + \frac{a/2}{\sqrt{k-a^2/4}} e^{-a/2t} \sin \sqrt{k-a^2/4}t$$

For  $k = .972$  and  $a = .825$

$$x_1(t) = e^{-.413t} \cos .896t + .461e^{-.413t} \sin .896t$$

Now find the equivalent phase margin model (see Appendix B). The phase margin model is simply the second order system with the same phase margin and crossover frequency as the given system. Therefore, the procedure will be to find the phase margin and crossover frequency (frequency at which  $|G(j\omega)| = 1$ ) of the given system, and then adjust the parameters of the second order system to give these same values.

The given system is  $G(j\omega) = \frac{10}{j\omega(j\omega+1)(j\omega+10)}$  and the phase margin and frequency when  $|G(j\omega)| = 1$  are  $\omega = .785$  and P.M. =  $\pi/2 - \tan^{-1}.919$ .

Now find the second order system of the form  $\frac{k}{j\omega(j\omega+a)}$  which crosses over at  $\omega = .785$  and at an angle of  $\theta = -\pi/2 - \tan^{-1}.919$

$$G_e(j\omega) = \frac{k}{j\omega(j\omega+a)} = \frac{k}{(a^2+\omega^2)^{1/2}} \angle(-\pi/2-\tan^{-1}\omega/a)$$

$$= 1 \angle(-\pi/2-\tan^{-1}.919)$$

$$\tan^{-1}\omega/a = \tan^{-1}.919$$

$$a = \frac{\omega}{.919} = \frac{.785}{.919} = .854$$

$$\frac{k}{.785(.73+.616)^{1/2}} = 1$$

$$k = .91$$

therefore,

$$G_e(s) = \frac{.91}{s(s+.854)}$$

The transient response to  $x_1(0) = 1$  is

$$x_1(t) = e^{-.427t} \cos .854t + .50e^{-.427t} \sin .854t$$

TABLE 3-1  
RESULTS OF EXAMPLE 3-1

System	System Response $x_1(t)$
Given System $\frac{10}{s(s+1)(s+10)}$	$1/92e^{-10.11t} + 91/92e^{-.445t} \cos.89t + .618e^{-.445t} \sin.89t$
Second Order Model $\frac{.972}{s(s+.825)}$	$e^{-.413t} \cos.896t + .461e^{-.413t} \sin.896t$
Phase Margin Model $\frac{.91}{s(s+.854)}$	$e^{-.427t} \cos.854t + .50e^{-.427t} \sin.854t$

TABLE 3-2  
RESULTS OF EXAMPLE 3-2

System	System Response $x_1(t)$
Given System $\frac{10}{s(s+1)(s+3)}$	$.249e^{-3.89t} + .851e^{-.055t} \cos 1.6t + .381e^{-.055t} \sin 1.6t$
Second Order Model $\frac{2.74}{s(s+.121)}$	$e^{-.061t} \cos 1.6t + .0381e^{-.061t} \sin 1.6t$
Phase Margin Model $\frac{2.5}{s(s+.125)}$	$e^{-.063t} \cos 1.58t + .039e^{-.063t} \sin 1.58t$

TABLE 3-3  
RESULTS OF EXAMPLE 3-3

System	System Response $x_1(t)$
Given System $\frac{1}{s(s+1)^2}$	$.178e^{-1.76t} + .822e^{-.12t}\cos.656t$ $+ .625e^{-.12t}\sin.656t$
Second Order Model $\frac{.557}{s(s+.249)}$	$e^{-.125t}\cos.736t + .170e^{-.125t}\sin.736t$
Phase Margin Model $\frac{.492}{s(s+.269)}$	$e^{-.135t}\cos.689t + .196e^{-.135t}\sin.689t$

TABLE 3-4  
RESULTS OF EXAMPLE 3-4

System	System Response $x_1(t)$
Given System $\frac{5}{s(s^2+5s+9)}$	$2.7e^{-t} - 1.7e^{-2t}\cos t - .7e^{-2t}\sin t$
Second Order Model $\frac{.817}{s(s+1.585)}$	$e^{-.793t}\cos.433t + 1.83e^{-.793t}\sin.433t$
Phase Margin Model $\frac{.937}{s(s+1.735)}$	$e^{-.867t}\cos.429t + 2.02e^{-.867t}\sin.429t$

TABLE 3-5  
RESULTS OF EXAMPLE 3-5

	System	System Response $x_1(t)$
Given System	$\frac{3}{s(s+2)(s+4)}$	$+ 3.94e^{-.7t} - 3.00e^{-t} + .06e^{-4.3t}$
Second Order Model	$\frac{.595}{s(s+1.24)}$	$e^{-.62t} \cos.308t + 2.01e^{-.62t} \sin.308t$
Phase Margin Model	$\frac{.502}{s(s+1.31)}$	$e^{-.66t} \cos.276t + 2.38e^{-.66t} \sin.276t$

$$G_S(s) = \frac{10}{s(s+1)(s+10)}$$

$$G_M(s) = \frac{.972}{s(s+.825)}$$

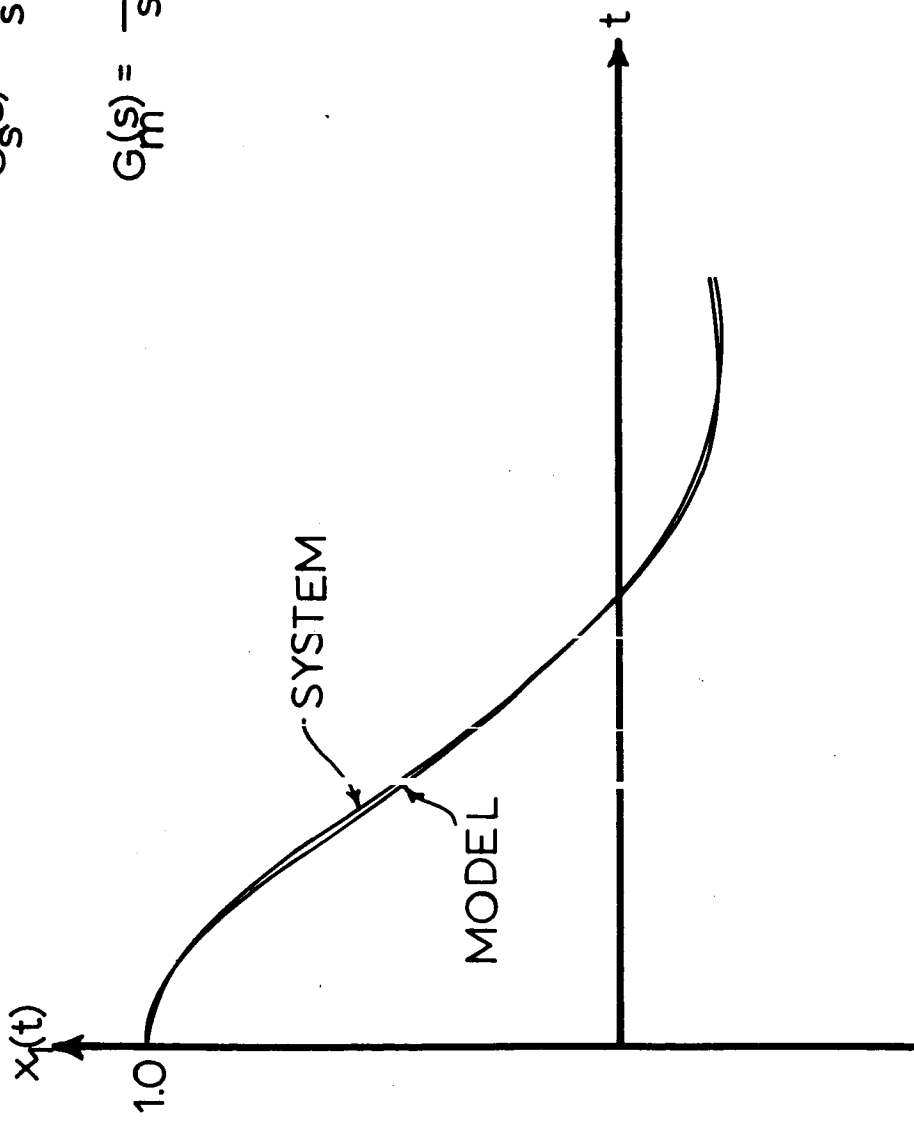


FIG. 3-7

TRANSIENT RESPONSES OF THE GIVEN SYSTEM AND MODEL FOR EXAMPLE 3-1



$$G(s) = \frac{10}{s(s+1)(s+3)}$$

$$G_m(s) = \frac{2.74}{s(s+1.21)}$$

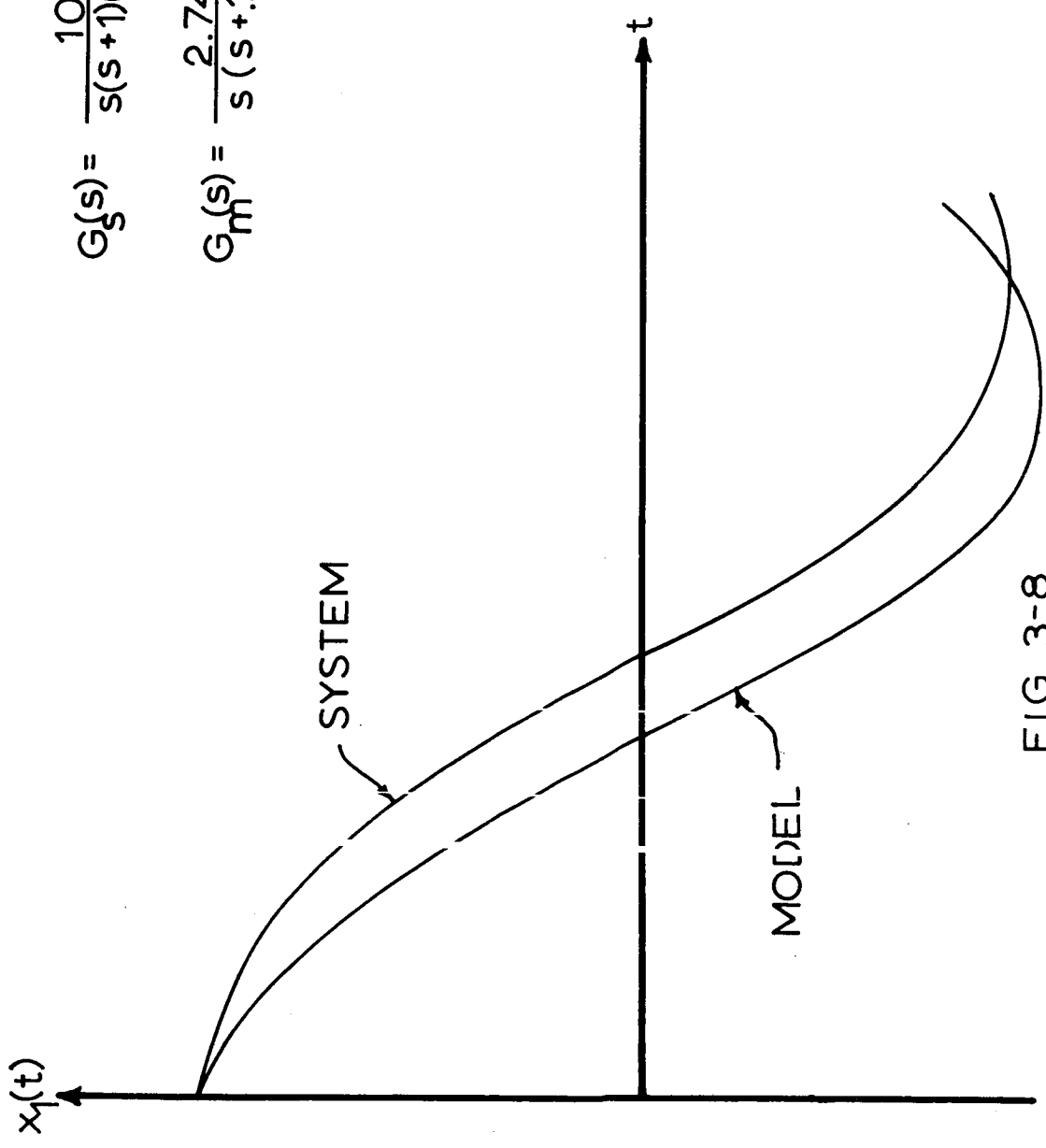


FIG. 3-8

TRANSIENT RESPONSES OF THE GIVEN SYSTEM AND MODEL FOR EXAMPLE 3-2

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

$$G_M(s) = \frac{.557}{s(s+.249)}$$

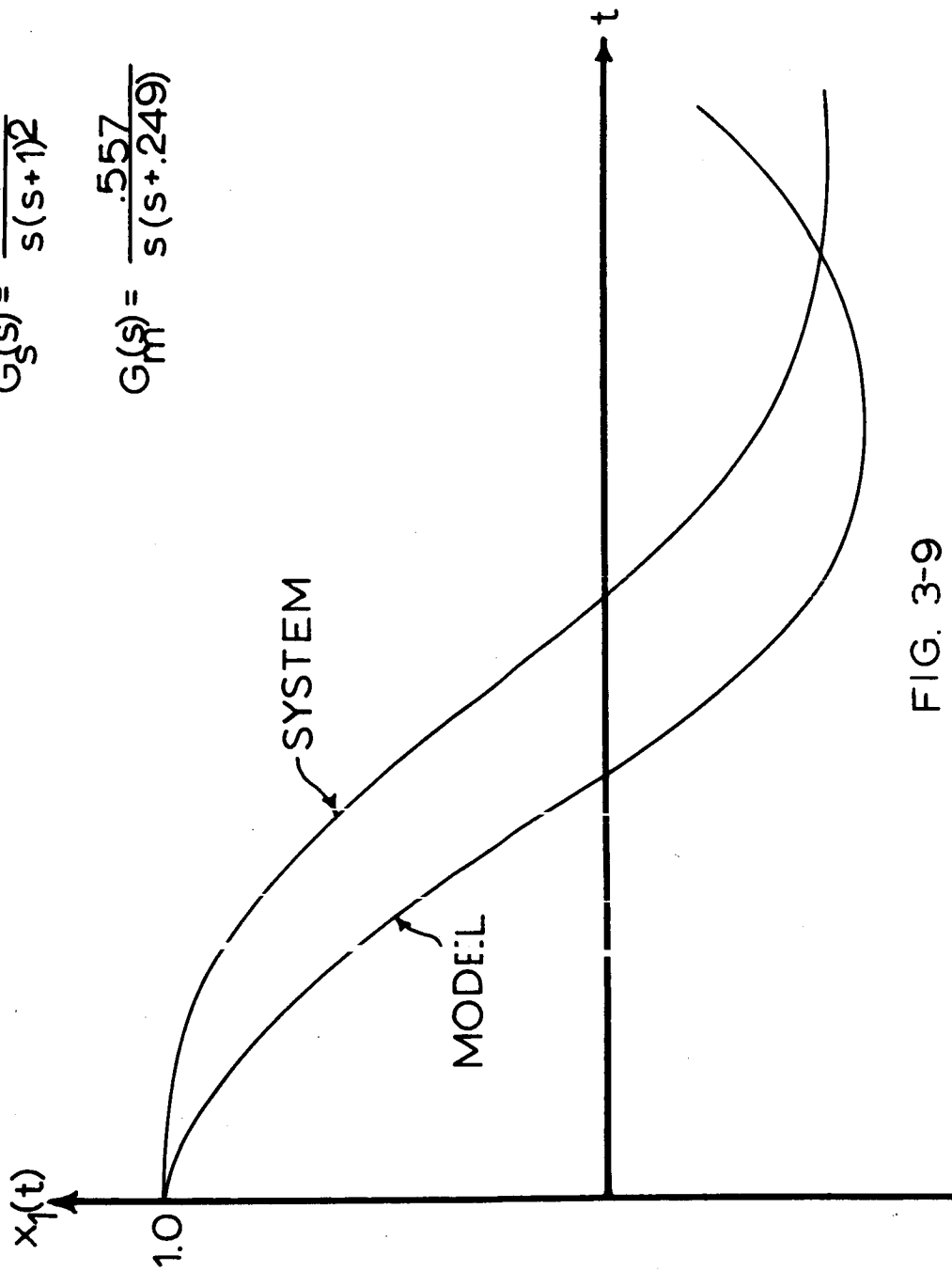
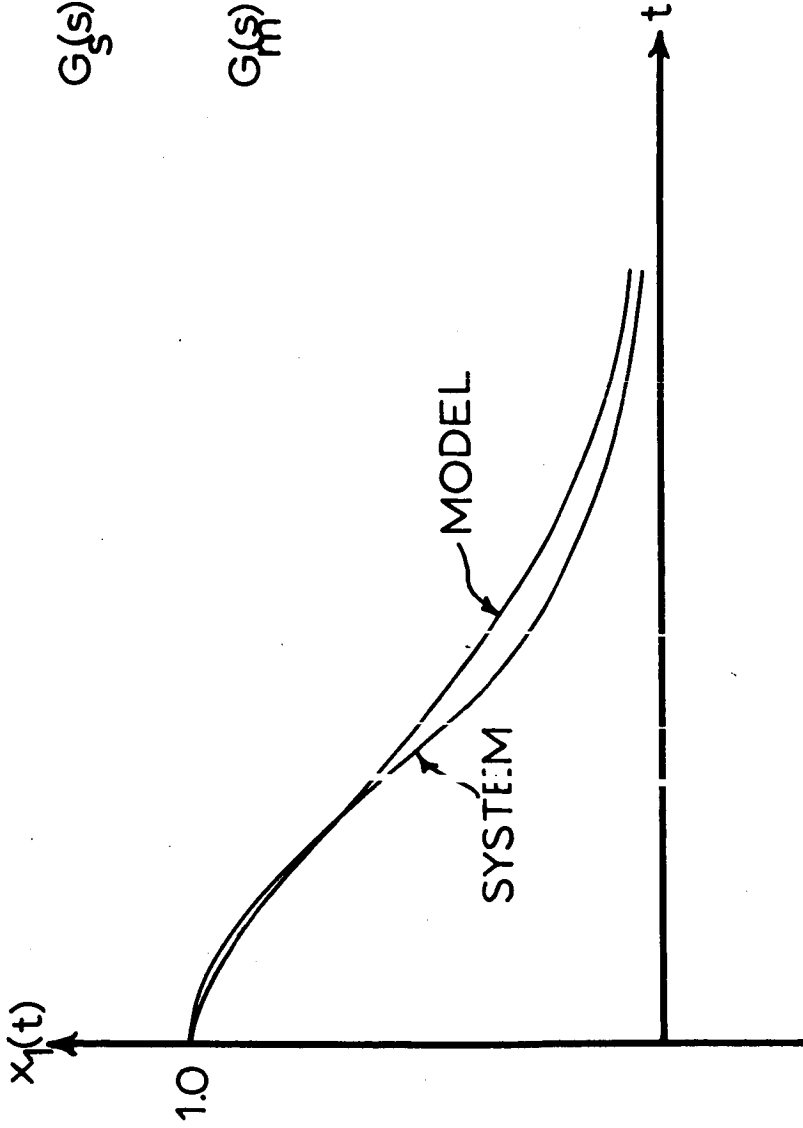


FIG. 3-9

TRANSIENT RESPONSES OF THE GIVEN SYSTEM AND MODEL FOR EXAMPLE 3-3



$$G_S(s) = \frac{5}{s(s^2 + 5s + 9)}$$

$$G_M(s) = \frac{.817}{s(s + 1.585)}$$

FIG. 3-10

TRANSIENT RESPONSES OF THE GIVEN SYSTEM AND MODEL FOR EXAMPLE 3-4

$$G(s) = \frac{3}{s(s+2)(s+4)}$$

$$G_M(s) = \frac{.595}{s(s+1.24)}$$

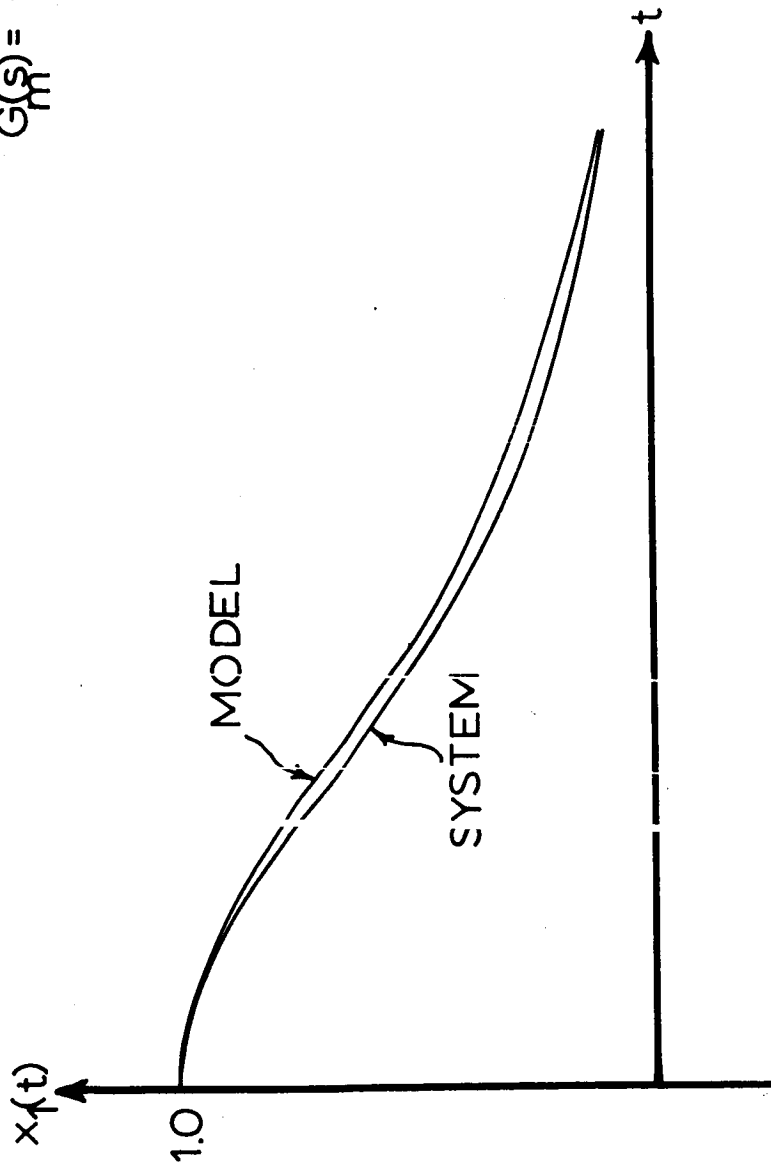


FIG. 3-11

TRANSIENT RESPONSES OF THE GIVEN  
SYSTEM AND MODEL FOR EXAMPLE 3-5

Upon studying the results in the tables it is immediately apparent that the phase margin and second order models are very similar for each example. This amazing fact makes the V surface modeling approach a very satisfactory one. Further insight into this similarity is given in Section 3.4. Since the models give nearly the same response, only the response of the second order model is compared to the response of the given system in Figs. 3-7 to 3-11.

Cases 1, 2, and 3 (Figs. 3-7, 3-8, 3-9) really belong in one main category. They are all underdamped third order systems whose real closed loop pole is further from the imaginary axis than the complex conjugate poles. See the root locus plots in Figs. 3-2, 3-3, 3-4. The only major difference between the three cases is the importance (relative distance from the imaginary axis) of this real pole. In Case 1 the pole is so far out from the imaginary axis that it has very little influence on the system behavior. Therefore, it seems likely that the closed loop poles of the model should closely coincide with the complex conjugate poles of the system. Such is the case and as expected the models give excellent approximations for this example. But as the pole moves in toward the imaginary axis it asserts greater influence on the system behavior and it is not so obvious where the closed loop poles of the model should be. Actually the closed

loop poles of the model move only slightly away from the system's complex conjugate poles as the real pole moves in toward the imaginary axis and because of the mounting influence of this real pole the models steadily decrease in accuracy as the pole achieves dominance (Figs. 3-8, 3-9). Upon looking at Figs. 3-7 to 3-9 it is seen that while the second order model has a similar frequency and damping factor, it becomes more and more out of phase with the actual system's response as the real pole achieves more dominance. This phase difference results in the model having a greater overshoot and a faster rise time than the third order system. It might be possible that this could be related to the angle of rotation  $\theta$  discussed previously.

In Case 4 (Example 3-4) the previous situation is reversed. Now the real pole is closer to the imaginary axis than the complex conjugate poles and therefore the real pole is the dominating pole. Looking at the root locus plot of this Case (Fig. 3-5) it is not at all obvious what form the closed loop poles of the model will take. Interestingly, the model is a slightly underdamped system. Upon inspection of Table 3-4 the second order and phase margin models do not seem to match the third order system at all, but upon plotting the responses in Fig. 3-9 it is seen that the second order model is a very good approximation to the given system.

Case 5 is concerned with third order systems with all real closed loop poles. Inspecting the root locus plot for this Case (Fig. 3-6) it seems that the closed loop poles of the model should be real and closely coincide with the dominant real closed loop poles of the system. It is surprising however that this is not the case at all. Both the second order and phase margin models give nearly the same underdamped system. Upon plotting the responses of the second order model and the given system in Fig. 3-11 it appears that this slightly underdamped system is a very good approximation to the actual system after all--an amazing result.

It should be apparent from these examples (especially Examples 3-1, 3-2, 3-3) that the accuracy of the model is greatly dependent on the relative positions of the closed loop poles of the given system. In systems of order greater than three the accuracy of the model is even more dependent on the location of all of the closed loop poles. An obvious way to increase the accuracy of the model to meet the demands of higher order systems is to give the model more closed loop poles by increasing the order of the model and, therefore, a third order model is developed in Chapter 4. Before proceeding with the third order model, however, it may be interesting to see why the second order model and the phase margin model are so similar.

### 3.4 Finding Phase Margin by Matching V Curves

The phase margin model was found by matching its phase margin to that of the given system (also its cross-over frequency). Interestingly enough, it turns out that one of the conditions determining the second order model (the model obtained by matching the axes products of the ellipses) fixes the phase margin of the model;  $|p_2^2| = a_1/4a_2^2$  uniquely determines the phase margin of the model and nothing else. This is true because for any second order system of the form  $\frac{a_1}{s(s+a_2)}$  the phase margin is given by  $a_1 = Ca_2^2$  where  $C = \cot\phi \sqrt{1 + \cot^2\phi}$  and  $\phi$  is the phase margin.

Proof:  $G_M^2(j\omega) = \frac{a_1}{j\omega(j\omega+a)}$

at crossover  $G_M^2(j\omega) = \frac{a_1}{j\omega(j\omega+a)} = 1 \angle (-\pi + \phi)$

( $\phi$  is the phase margin)

$$\frac{a_1}{(\omega^2 + a_2^2)^{1/2}} \angle (-\pi/2 - \tan^{-1}\omega/a_2) = 1 \angle (-\pi + \phi)$$

$$-\pi/2 - \tan^{-1}\omega/a_2 = -\pi + \phi$$

$$\tan^{-1}\omega/a_2 = \pi/2 - \phi$$

$$\omega/a_2 = \tan(\pi/2 - \phi) = \cot\phi$$

$$\omega = a_2 \cot\phi$$



$$a_1 = \omega \sqrt{\omega^2 + a_2^2}$$

$$a_1 = a_2 \cot \phi \sqrt{a_2^2 \cot^2 \phi + a_2^2}$$

$$a_1 = a_2^2 \cot \phi \sqrt{1 + \cot^2 \phi}$$

or  $a_1 = C a_2^2$  where  $C = \cot \phi \sqrt{1 + \cot^2 \phi}$

Therefore,  $|\underline{P}_2^2| = a_1 / 4a_2^2 = C a_2^2 / 4a_2^2 = \frac{\cot \phi \sqrt{1 + \cot^2 \phi}}{4}$

and  $|\underline{P}_2^2|$  completely specifies the phase margin of the model. Since the phase margin model and the second order model are similar it follows that  $|\underline{P}_2^2|$  must give a good indication of the phase margin of the nth order system.

Example (3-6) Referring to Example 3-1 in Section 3.3 note that the actual phase margin of the given system is

$$\phi = \pi + \theta = \pi/2 - \tan^{-1} .919$$

$$\phi = 90^\circ - 42.6^\circ = 47.4^\circ$$

$$\phi = 47.4^\circ$$

Now using  $|\underline{P}_2^2|$  to find the phase margin of the second order model (Example 3-1)

$$\cot \phi \sqrt{1 + \cot^2 \phi} = 4 |\underline{P}_2^2| = 4 \left( \frac{1.43}{4} \right) = 1.43$$

$$\cot^2 \phi (1 + \cot^2 \phi) = 2.05$$

$$\cot^2 \phi = 1.007$$

$$\cot\phi = 1.003$$

$$\phi = 45.1^\circ$$

$45.1^\circ$  is quite close to  $47.4^\circ$  and thus  $|P_{-2}^2|$  does seem to give an accurate indication of the phase margin of the given system.

## CHAPTER 4

### THE THIRD ORDER MODEL

#### 4.1 Development of the Model

As seen from the results of Chapter 3, the second order model and the phase margin model give much the same results. In fact, it appears that the second order model is approximately equivalent to the phase margin model. While the models are good approximations for many systems, in some cases more accuracy (a better model) may be desired. Obviously, a better model would be a third order model, and although it would be harder to analyze than a second order model, it should certainly give better results. Third order models find little use in modeling third order systems since the best third order model is the given third order system itself (second order models must suffice). But, third order models are practical for higher order systems where second order models may not give accurate results. Models of order higher than three are questionable, since then the analysis of the model would become too difficult for the possible gain in accuracy.

It is tempting to extend the latter method for finding second order models (the method of matching the products of the major axes) to third order models. The

V surface (ellipsoid) of the third order model is given as

$V = \underline{x}^T \underline{M} \underline{x}$  where,

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} ; \quad \underline{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{12} & m_{22} & m_{23} \\ m_{13} & m_{23} & m_{33} \end{bmatrix}$$

The V curve of the nth order system is given by  $V = \underline{x}^T \underline{P} \underline{x}$

where,

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} ; \quad \underline{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \cdot & \cdot & \cdot & P_{1n} \\ P_{12} & P_{22} & P_{23} & & & & \cdot \\ P_{13} & P_{23} & P_{33} & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ P_{1n} & \cdot & \cdot & \cdot & \cdot & \cdot & P_{nn} \end{bmatrix}$$

The V curve intersection of the nth order system (inter-  
section of the nth order V curve in three space--

ellipsoid) is given by  $V = \underline{x}^{3T} \underline{P}^3 \underline{x}^3$  where  $\underline{x}^3$  is a vector  
of the first three components of  $\underline{x}$  and  $\underline{P}^3$  is the upper left  
hand 3 by 3 submatrix of  $\underline{P}$ .

$$\underline{x}^3 = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} ; \quad \underline{P}^3 = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{12} & P_{22} & P_{23} \\ P_{13} & P_{23} & P_{33} \end{bmatrix}$$

To match the axes products for a particular  $\dot{V}$ , it is necessary and sufficient, by extending the reasoning in Section 3.2 to third order, that  $|\underline{M}| = |\underline{P}^3|$ . To completely specify the model it is necessary to satisfy three such relationships for three different choices of  $\dot{V}$ . As before, a convenient choice for these  $\dot{V}$ 's is  $\dot{V}_1 = -x_1^2$ ,  $\dot{V}_2 = -x_2^2$ , and  $\dot{V}_3 = -x_3^2$ . The expressions for  $\underline{M}_i$  ( $i = 1, 2, 3$ ) can be obtained from part 2 of Appendix A or from  $\underline{A}^T \underline{M}_i + \underline{M}_i \underline{A} = -\underline{Q}_i$  where  $\underline{A}$  is the system matrix of the third order model.

$$\underline{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_1 & -a_2 & -a_3 \end{bmatrix}, \quad \underline{Q}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\underline{Q}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \underline{Q}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\text{Then, } |\underline{M}_1| = \frac{1/a_1}{8(a_2 a_3 - a_1)^2}$$

$$|\underline{M}_2| = \frac{a_1}{8(a_2 a_3 - a_1)^2}$$

$$|\underline{M}_3| = \frac{a_1^3}{8(a_2 a_3 - a_1)^2}$$

Or,  $|\underline{M}_1| |\underline{M}_3| = |\underline{M}_2|^2$ ; but this condition is not met in general by higher order systems. It is, therefore, impossible to match the axes products of the curves for

these three choices of  $\dot{V}$ . Another  $\dot{V}$  must be chosen so that no conflicting requirement results. At present this choice is not apparent and therefore the first method is used for third order models (matching  $V$  curves for a particular  $\dot{V}$ ).

The process for matching  $V$  surfaces in three space (ellipsoids) is exactly the same as for two space. Again, it is impossible to match the two  $V$  surfaces term by term, since there are six such terms to match ( $V = p_{11}x_1^2 + 2p_{12}x_1x_2 + 2p_{13}x_1x_3 + p_{22}x_2^2 + 2p_{23}x_2x_3 + p_{33}x_3^2$ ) with only three variable parameters. Therefore, the coordinate axes must be rotated for both  $V$  surfaces, to eliminate the cross terms leaving only diagonal elements. The diagonal terms of the two transformed matrices may then be set equal to determine the three parameters of the third order model. This rotation is even harder to accomplish in the third order case than the second order case, because there are now three cross terms to eliminate instead of only one. Also, the elements of the  $\underline{M}$  matrix ( $V = \underline{x}^T \underline{M} \underline{x}$  is the  $V$  curve of the third order model) are not constants but functions of the three variable system parameters and, therefore, the matching of diagonal terms must be a procedure of trial and error--not a very satisfactory result. An attempt is thus made as in the second order case to find a  $\dot{V}$  that immediately gives  $\underline{M}$  in diagonal form.

Given  $\dot{V} = -\underline{x}^T \underline{Q} \underline{x}$ , try to find  $\underline{Q}$  so that  $V = \underline{x}^T \underline{M} \underline{x}$  gives  $\underline{M}$  in diagonal form.  $\underline{Q}$  and  $\underline{M}$  are related by the matrix equation  $\underline{A}^T \underline{M} + \underline{M} \underline{A} = -\underline{Q}$  where  $\underline{A}$  is the system matrix.

$$\underline{Q} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{12} & q_{22} & q_{23} \\ q_{13} & q_{23} & q_{33} \end{bmatrix}, \quad \underline{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{12} & m_{22} & m_{23} \\ m_{13} & m_{23} & m_{33} \end{bmatrix}, \quad \underline{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_1 & -a_2 & -a_3 \end{bmatrix}$$

Equating the elements of the matrices  $(\underline{A}^T \underline{M} + \underline{M} \underline{A})$  and  $-\underline{Q}$ , the following six equations result.

1.  $-2a_1 m_{13} = -q_{11}$
2.  $-a_1 m_{23} + m_{11} - a_2 m_{13} = -q_{12}$
3.  $-a_1 m_{33} + m_{12} - a_3 m_{13} = -q_{13}$
4.  $2(m_{12} - a_2 m_{23}) = -q_{22}$
5.  $m_{13} - a_2 m_{33} + m_{22} - a_3 m_{23} = -q_{23}$
6.  $2(m_{23} - a_3 m_{33}) = -q_{33}$

Solving these simultaneous equations for the elements of  $\underline{M}$  in terms of the "q's" and the "a's," the off diagonal elements of  $\underline{M}$  turn out to be:

$$m_{13} = \frac{q_{11}}{2a_1}$$

$$m_{12} = \frac{a_2 a_3^2 q_{11} + a_1^2 q_{22} + a_1^2 a_2 q_{33} - 2a_1 a_2 a_3 q_{13}}{2a_1 (a_3 a_2 - a_1)}$$

$$m_{23} = \frac{a_3^2 q_{11} + a_1 a_3 q_{22} + a_1^2 q_{33} - 2a_1 a_3 q_{13}}{2a_1(a_3 a_2 - a_1)}$$

The only way to make  $m_{13}$  zero, is to set  $q_{11} = 0$ . To force  $m_{12}$  and  $m_{23}$  to be zero, it is necessary to satisfy

$$1. \quad a_2 a_3^2 q_{11} + a_1^2 q_{22} + a_1^2 a_2 q_{33} = 2a_1 a_2 a_3 q_{13}$$

$$2. \quad a_3^2 q_{11} + a_1 a_3 q_{22} + a_1^2 q_{33} = 2a_1 a_3 q_{13}$$

$$3. \quad 2a_1(a_3 a_2 - a_1) > 0$$

$$4. \quad \underline{Q} \text{ positive semi definite or positive definite}$$

If it is possible to satisfy these conditions,  $q_{13}$ ,  $q_{22}$ ,  $q_{33}$  must be functions of  $a_1$ ,  $a_2$ ,  $a_3$ . In other words,  $\underline{Q}$  is a function of the parameters of the model--an unfortunate result since these parameters are unknown. Therefore, the procedure of matching the  $V$  surfaces for this  $\dot{V} = -\underline{x}^T \underline{Q} \underline{x}$  must again be a process of trial and error with the preliminary step being a guess of  $\dot{V}$  or  $\underline{Q}$ . It is apparent that  $\underline{Q}$  must have elements which are known constants and not functions of  $a_1$ ,  $a_2$ ,  $a_3$ . With such a  $\underline{Q}$ ,  $m_{12}$  and  $m_{13}$  can never be made zero.  $m_{13}$  can be made zero however by setting  $q_{11} = 0$ . Two such  $\underline{Q}$ 's are



$$\underline{Q}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \underline{Q}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

or  $\dot{V} = -x_2^2$  and  $\dot{V} = -x_3^2$ .

The resulting  $\underline{M}$  must still be transformed to its diagonal form. Rather than doing this an approximation scheme is used to match the two  $V$  surfaces: choose  $\dot{V}$  so as to eliminate as many cross terms as possible; transform the  $\underline{P}^3$  matrix ( $V = \underline{x}^{3T} \underline{P}^3 \underline{x}^3$  is the  $V$  surface intersection in three space of the  $n$ th order system) to eliminate the same cross-terms that are missing in the  $\underline{M}$  matrix (in a sense, these cross terms have now been matched); force the diagonal elements of  $\underline{M}$  and the transformed  $\underline{P}^3$  matrix to be equal ignoring the remaining cross terms. This approximation is shown to be good or bad by regarding the remaining cross terms now that the diagonals are matched. If the cross terms are similar in value the approximation is a good one. If the cross terms are not at all similar then the approximation is not good. At the very least this approximation procedure gives a starting place for the trial and error process of matching the shapes of the  $V$  surfaces exactly. Usually, however, if the resulting third order model is to be any good at all, the  $V$  surfaces  $\underline{x}^T \underline{M} \underline{x}$  and  $\underline{x}^{3T} \underline{P}^3 \underline{x}^3$  should be closely oriented to begin with, and therefore the cross terms should closely match when

diagonal terms are made equal. Conversely, if the  $V$  surfaces are not oriented in the same direction when their shapes are exactly matched, then the model will not be a good representation of the system anyway and  $\underline{P}^3$  and  $\underline{M}$  will not be similar. In other words, the approximation will be bad but the exact model would have been bad anyway. Thus the approximation process is used to find the third order model. As stated before, the choices of  $\dot{V}$  which eliminate the most cross terms of  $\underline{M}$  are  $\dot{V} = -x_2^2$  and  $\dot{V} = -x_3^2$ . Both choices for  $\dot{V}$  result in the  $x_1x_3$  term being zero. It is necessary, therefore, to transform  $\underline{x}^{3T} \underline{P}^3 \underline{x}^3$  to eliminate the  $x_1x_3$  term. Recalling the procedure for eliminating cross terms in Section 3.1, a suitable change of variables is:

$$y_1 = \cos\theta x_1 + \sin\theta x_3$$

$$y_2 = x_2$$

$$y_3 = -\sin\theta x_1 + \cos\theta x_3$$

$$\text{or, } \underline{y} = \underline{B}\underline{x}^3, \quad \underline{B} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$\underline{B}^{-1} = \underline{B}^T = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$\text{therefore, } \underline{x}^3 = \underline{B}^T \underline{y}$$

$$\text{and } \underline{V} = \underline{x}^3 \underline{P}^3 \underline{x}^3 = \underline{y}^T \underline{B} \underline{P}^3 \underline{B}^T \underline{y} = \underline{y}^T \underline{R} \underline{y}$$

$$\underline{R} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{12} & p_{22} & p_{23} \\ p_{13} & p_{23} & p_{33} \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$\underline{R} = \begin{bmatrix} (p_{11}\cos^2\theta + 2p_{13}\cos\theta\sin\theta + p_{33}\sin^2\theta) \\ (p_{12}\cos\theta + p_{23}\sin\theta) \\ (\sin\theta\cos\theta(p_{33}-p_{11}) + p_{13}(\cos^2\theta - \sin^2\theta)) \end{bmatrix}$$

$$(p_{12}\cos\theta + p_{23}\sin\theta)$$

$$p_{22}$$

$$(-p_{12}\sin\theta + p_{23}\cos\theta)$$

$$(\sin\theta\cos\theta(p_{33}-p_{11}) + p_{13}(\cos^2\theta - \sin^2\theta))$$

$$(-p_{12}\sin\theta + p_{23}\cos\theta)$$

$$(p_{11}\sin^2\theta - 2p_{13}\cos\theta\sin\theta + p_{33}\cos^2\theta)$$

But  $r_{13}$  must be zero, therefore,

$$r_{13} = \sin\theta\cos\theta(p_{33}-p_{11}) + p_{13}(\cos^2\theta - \sin^2\theta) = 0$$

$$1/2(p_{11}-p_{33})\sin 2\theta = p_{13}\cos 2\theta$$

$$\tan 2\theta = \frac{2p_{13}}{p_{11} - p_{33}}$$

$$\theta = 1/2 \tan^{-1} \left( \frac{2p_{13}}{p_{11} - p_{33}} \right) \quad (4-1)$$

The other terms of the R matrix are:

$$r_{11} = p_{11} \cos^2 \theta + 2p_{13} \cos \theta \sin \theta + p_{33} \sin^2 \theta$$

$$r_{12} = p_{12} \cos \theta + p_{23} \sin \theta$$

$$r_{22} = p_{22} \quad (4-2)$$

$$r_{23} = -p_{12} \sin \theta + p_{23} \cos \theta$$

$$r_{33} = p_{11} \sin^2 \theta - 2p_{13} \cos \theta \sin \theta + p_{33} \cos^2 \theta$$

Using Eqs. (4-1) and (4-2) it is possible to transform the P<sup>3</sup> matrix to R to eliminate the  $p_{13}$  term. Since  $\dot{V} = -x_2^2$  yields an M matrix with simpler expressions for the diagonal terms than does  $\dot{V} = -x_3^2$ ,  $\dot{V} = -x_2^2$  is used as a basis to find the model. From the tables in part 2 of Appendix A for  $\dot{V} = -x_2^2$ , M becomes:

$$\underline{M} = \frac{1}{2(a_3 a_2 - a_1)} \begin{bmatrix} a_3 a_1 & a_1 & 0 \\ a_1 & (a_3^2 + a_2) & a_3 \\ 0 & a_3 & 1 \end{bmatrix}$$

where  $a_1, a_2, a_3$  are the variables of the model. Equating the diagonal terms of R and M,

$$\frac{a_3 a_1}{2(a_3 a_2 - a_1)} = r_{11}$$

$$\frac{a_3^2 + a_2}{2(a_3 a_2 - a_1)} = r_{22}$$

$$\frac{1}{2(a_3 a_2 - a_1)} = r_{33}$$

Solving these three expressions for  $a_1$ ,  $a_2$ ,  $a_3$ ,

$$a_2 = r_{22}/r_{33} - a_3^2$$

$$a_1 = r_{11}/(a_3 r_{33}) \quad (4-3)$$

$$(r_{22}/r_{33})a_3^2 - a_3^4 - (1/2r_{33})a_3 = r_{11}/r_{33}$$

The step by step procedure for determining the third order model is:

1. Pick  $\dot{\underline{v}} = -\underline{x}_2^2 = -\underline{x}^T \underline{Q}_2 \underline{x}$  and solve for  $\underline{P}$ , an  $n$  by  $n$  matrix, from the matrix equation  $\underline{A}'^T \underline{P} + \underline{P} \underline{A}' = -\underline{Q}_2$  where  $\underline{A}'$  is the  $n$ th order system matrix (use tables in part 3 of the appendix for fourth order system).
2. Take the upper left hand 3 by 3 submatrix of  $\underline{P}$  for  $\underline{P}^3$ .
3. Rotate the coordinate axes of the state space to eliminate the  $p_{13}$  term.
4. Solve for  $a_1$ ,  $a_2$ ,  $a_3$  (parameters of the third order model) from the Eqs. (4-3). The model is now completely specified by,

$$\underline{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_1 & -a_2 & -a_3 \end{bmatrix}$$

Obviously, the work needed to find the third order model is much greater than that needed for the second order model; the  $\underline{P}^3$  matrix must be transformed, and then a non-linear algebraic equation must be solved to find  $a_3$ . Hopefully, the result of this additional labor is a better model.

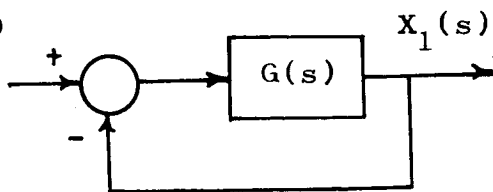
#### 4.2 Examples of the Third Order Model

In this section, two fourth order systems are modeled with the third order model. The first example is worked in detail and the results are given for the second. This model is compared with the second order and phase margin models in Tables 4-1 and 4-2.

##### Example 4-1 (Finding the third order model for a fourth order system)

Given the fourth order system  $G(s)$  find the third order model and compare its transient response to that of the system for  $x_1(0) = 1$ .

$R = 0$



$$G(s) = \frac{6}{s(s+1)(s+2)(s+4)}$$

First find the system matrix  $\underline{A}'$  from the block diagram.

$$\frac{X_1(s)}{-X_1(s)} = \frac{6}{s^4 + 7s^3 + 14s^2 + 8s + 6}$$

$$\overset{\cdot\cdot\cdot}{x} + 7\overset{\cdot\cdot}{x} + 14\overset{\cdot}{x} + 8x = -6x$$

Defining

$$x = x_1$$

$$\dot{x} = x_1 = x_2$$

$$\ddot{x} = x_2 = x_3$$

$$\overset{\cdot\cdot\cdot}{x} = x_3 = x_4$$

then the differential equation can be written as,

$$\dot{x}_4 = -6x_1 - 8x_2 - 14x_3 - 7x_4$$

Then the system matrix  $\underline{A}'$  becomes

$$\underline{A}' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -6 & -8 & -14 & -7 \end{bmatrix}$$

The system transient response  $x_1(t)$  for  $x_1(0) = 1$ ,  $x_2(0) = x_3(0) = x_4(0) = 0_x$  is:

$$x_1(t) = .31e^{-3t} - .16e^{-3.6t} + .85e^{-2t} \cos.72t \\ + .74e^{-.2t} \sin.72t$$

Now find the second order model and its transient response for  $x_1(0) = 1$  (refer to the procedure at the end of Section 4.1).

1. Pick  $\dot{V} = -x_2^2 = -\underline{x}^T \underline{Q} \underline{x}$ , then for the fourth order system

$$\underline{Q}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Using part 3 of Appendix A and the system matrix  $\underline{A}$  ( $a_1 = 6$ ,  $a_2 = 8$ ,  $a_3 = 14$ ,  $a_4 = 7$ )  $\underline{P}_2$  becomes

$$\underline{P}_2 = \frac{1}{2(426)} \begin{bmatrix} 540 & 294 & 42 & ( ) \\ 294 & 1610 & 686 & ( ) \\ 42 & 686 & 351 & ( ) \\ ( ) & ( ) & ( ) & ( ) \end{bmatrix}$$

2. Take the upper left hand 3 by 3 submatrix of  $\underline{P}_2$  for  $\underline{P}_2^3$

$$\underline{P}_2^3 = \frac{1}{2(426)} \begin{bmatrix} 540 & 294 & 42 \\ 294 & 1610 & 686 \\ 42 & 686 & 351 \end{bmatrix}$$

3. Eliminate the  $p_{13}$  term by rotation of axes using Eqs. (4-1) and (4-2).  $\underline{R}$  becomes,



$$\underline{R} = \frac{1}{852} \begin{bmatrix} 548 & 431 & 0 \\ 431 & 1610 & 610 \\ 0 & 610 & 342 \end{bmatrix} = \begin{bmatrix} .643 & .505 & 0 \\ .505 & 1.890 & .716 \\ 0 & .716 & .402 \end{bmatrix}$$

4. Now solve for  $a_1$ ,  $a_2$ ,  $a_3$  (variable parameters of the model) from Eqs. (4-3).

$$a_1 = \frac{1.605}{a_3}$$

$$a_2 = 4.71 - a_3^2$$

$$4.71a_3^2 - a_3^4 - 1.244a_3 = 1.605$$

Solving the nonlinear equation,

$$\begin{array}{l} a_3 = 1.90 \\ a_2 = 1.10 \\ a_1 = .845 \end{array} , \underline{M} = \begin{bmatrix} .643 & .340 & 0 \\ .340 & 1.890 & .764 \\ 0 & .764 & .402 \end{bmatrix}$$

Actually there are two real roots for  $a_3$ , but  $a_3 = 1.90$  gives an  $\underline{M}$  matrix whose off diagonal elements most closely match those of  $\underline{P}_2^3$ . The third order model becomes,

$$\underline{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -.845 & -1.10 & -1.90 \end{bmatrix}$$

Choosing a model in the form  $G_M^3(s) = \frac{k}{s(s^2 + as + b)}$ ,

$$\text{then } k = a_1 = .845$$

$$a = a_3 = 1.90$$

$$b = a_2 = 1.10$$

$$\text{and } G_M^3(s) = \frac{.845}{s(s^2 + 1.90s + 1.10)}$$

For the initial condition  $x_1(0) = 1$ ,  $x_1(t)$  becomes,

$$x_1(t) = .232e^{-1.54t} + .768e^{-.182t} \cos .72t \\ + .68e^{-.182t} \sin .72t$$

To find the second order model and the phase margin model the procedure is exactly the same as in Example 3-1, and therefore is not repeated here.

From the results in Tables 4-1 and 4-2 notice that, as with third order systems, the phase margin and second order models are almost the same for both examples; they each give the same fair approximation. The third order model, however, gives a very good approximation to the given system as seen from Figs. 4-1 and 4-2, and perhaps this is the best justification of all for the approximate method used in matching the V surfaces. Looking at the figures it appears that while the second order model can give the same frequency and damping factor as the given system, its response lags the system response. The third

order model, however, can give the same frequency, damping factor, and also the same phase as the system--a result of three adjustable parameters instead of two.

This concludes the discussion of the V surface modeling technique. The final chapter summarizes the method and its results and suggests further areas for further research.

TABLE 4-1  
RESULTS OF EXAMPLE 4-1

	System	System Response $x_1(t)$
Given System	$\frac{6}{s(s+1)(s+2)(s+4)}$	$.31e^{-3t} - .16e^{-3.6t} + .85e^{-.2t} \cos.72t + .74e^{-.2t} \sin.72t$
Third Order Model	$\frac{.845}{s(s^2+1.90s+1.10)}$	$.232e^{-1.54t} + .768e^{-.182t} \cos.72t + .68e^{-.182t} \sin.72t$
Second Order Model	$\frac{.533}{s(s+.351)}$	$e^{-.176t} \cos.71t + .249e^{-.176t} \sin.71t$
Phase Margin Model	$\frac{.446}{s(s+.395)}$	$e^{-.198t} \cos.638t + .31e^{-.198t} \sin.638t$

TABLE 4-2  
RESULTS OF EXAMPLE 4-2

	System	System Response $x_1(t)$
Given System	$\frac{10}{s(s+1)^2(s+10)}$	$.307e^{-1.82t} - .0e^{-9.99t} + .693e^{-.095t} \cdot \cos.738t + .533e^{-.095t} \sin.738t$
Third Order Model	$\frac{.832}{s(s^2+1.71s+.82)}$	$.206e^{-1.53t} + .794e^{-.09t} \cos.733t + .615e^{-.09t} \sin.733t$
Second Order Model	$\frac{.533}{s(s+.197)}$	$e^{-.099t} \cos.725t + .136e^{-.099t} \cdot \sin.725t$
Phase Margin Model	$\frac{.485}{s(s+.216)}$	$e^{-.108t} \cos.688t + .157e^{-.108t} \cdot \sin.688t$

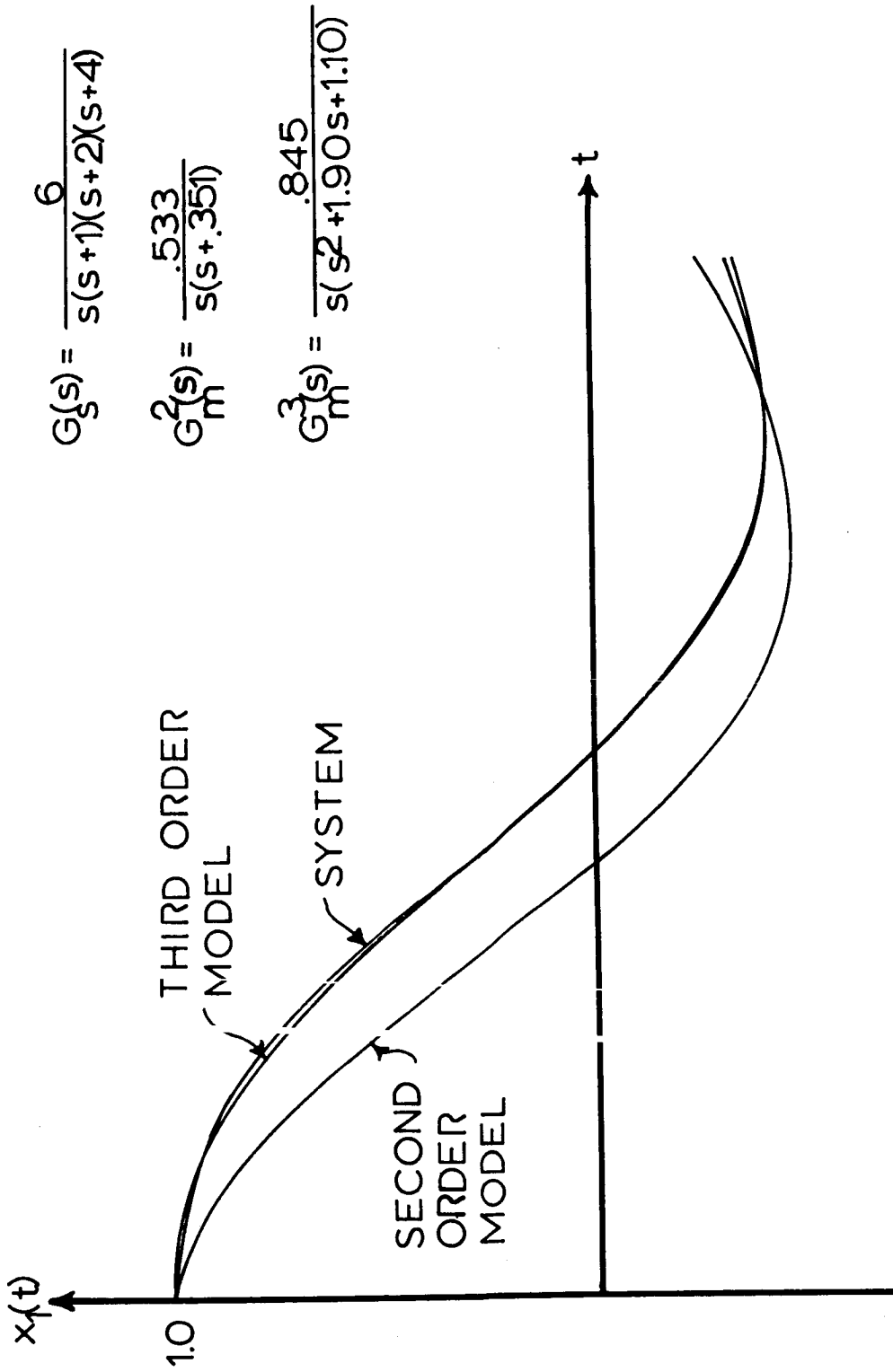


FIG. 4-1

TRANSIENT RESPONSES OF THE GIVEN SYSTEM AND MODELS FOR EXAMPLE 4-1

$$G(s) = \frac{10}{s(s+1)^2(s+10)}$$

$$G_m^2(s) = \frac{.535}{s(s+.197)}$$

$$G_m^3(s) = \frac{.832}{s(s^2+1.71s+.82)}$$

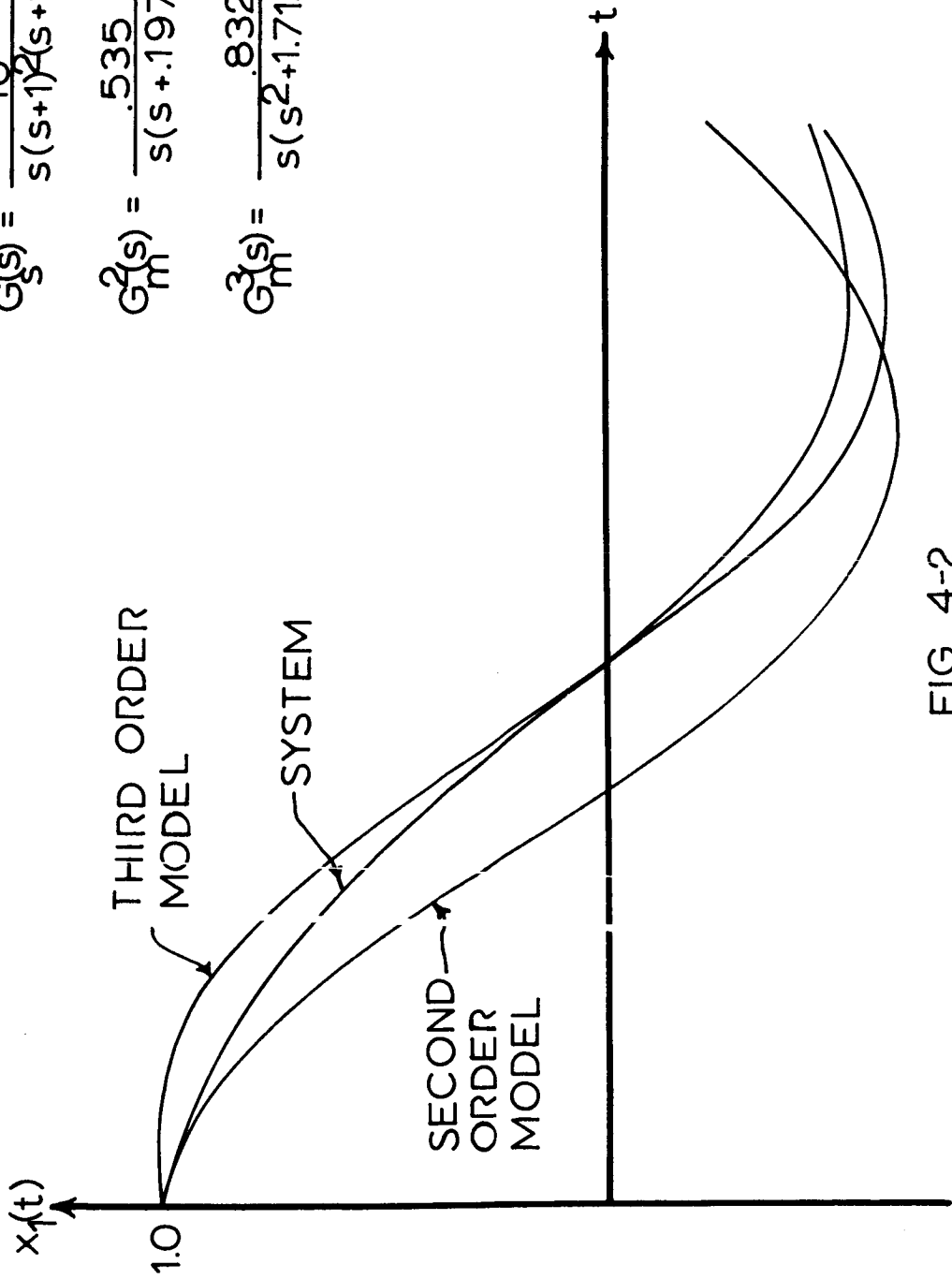


FIG. 4-2

TRANSIENT RESPONSES OF THE GIVEN SYSTEM AND MODELS FOR EXAMPLE 4-2

## CHAPTER 5

### SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH

#### 5.1 Summary

In this paper, the Second Method of Liapunov was used to develop a method for obtaining a model for high order control systems. This was accomplished by matching the surfaces described in the state space by the Liapunov functions of the model and the system. In particular, the second and third order models were found for systems without zeros and shown to be good approximations to the given systems.

The second order model was seen to be very similar to the model found from phase margin techniques. The second order model, however, was obtained directly from time domain considerations (Liapunov  $V$  curves in state space) and may be more appealing in that sense. Also this method of modeling is easily programmed on a digital computer although the model can be found by hand calculation using the tables in Appendix A for up to fourth order systems.

Of further interest, a third order model was obtained using these techniques which gave an even better approximation to the given system than either the second



order or the phase margin model. Higher order models are of course possible, but their usefulness in analysis may be limited by their own complexity.

As with any approximate analysis procedure, it is desirable to have some means of determining the accuracy of the analysis and often this is not possible except by direct comparison with the actual solution of the system. With the V surface modeling technique, however, one indication of its accuracy is how closely the V surfaces are oriented when their shapes are matched.

## 5.2 Suggestions for Further Research

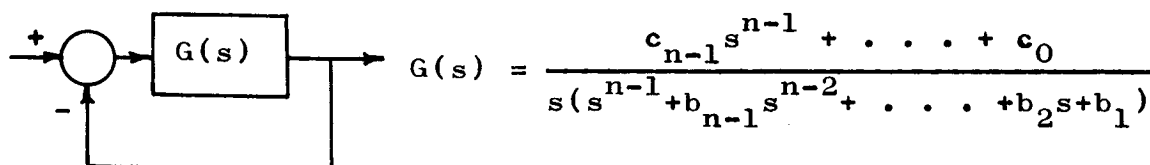
This thesis exists to introduce an approach which as of now appears promising. The paper merely points out the fact that the behavior of a control system is related to the shape and orientation of its V surface. It is not meant to be a final conclusive answer to the problem of modeling and much work yet remains to be done, some of which is suggested below.

First of all there is the problem of systems with zeros. This problem exists because the phase variable formulation of the system matrix A is not unique when zeros are allowed.<sup>2</sup> As seen from Eq. (2-2), the use of phase variables results in an A matrix which is dependent only on the characteristic equation of the system. When the types of systems considered are limited to the form of

Eq. (2-1), then the characteristic equation uniquely determines the system. But when the systems can have zeros, the characteristic equation does not uniquely determine the system; i.e., the A matrix can represent many different systems with the same characteristic equation. The use of phase variables still results in a valid model, but difficulty arises in interpreting the model's A matrix. Since the matrix defines only the characteristic equation of the model, it is impossible to know whether the model has a zero or the value of this zero. To make the A matrix unique a new set of variables must be used to describe the system. These variables must obey certain rules. Of course they must yield an A matrix which is unique for any type of system. Also, for the matching of V surfaces to have any meaning, the space containing these surfaces must be identical for the model and the system. In other words, the first k variables of the system must be identical to the k variables of the model, where k is the order of the model. Finally, the method of choosing the n variables of an nth order system should be consistent for all types of systems. If such a set of variables exists, then formulas for the model can be developed as in Chapters 3 and 4.

As mentioned previously, when the modeling method is dependent on phase variables, it is still valid even when the system to be analyzed contains zeros in its open

loop transfer function. The resulting A matrix is valid but it specifies only the characteristic equation of the model. By some additional work it may also be possible to determine the exact form of the model. One way which seems to give reliable results is to observe the effect that the zeros have on the characteristic equation. For an  $n$ th order system of the form



the characteristic equation is

$$s^n + (b_{n-1} + c_{n-1})s^{n-1} + \dots + s^2(b_2 + c_2) + s^1(b_1 + c_1) + c_0 = 0$$

When  $G(s)$  has no zeros, then only  $c_0$  appears in the characteristic equation. When one zero is present, then  $c_1$  is present in the  $s$  term. For two zeros,  $c_2$  is present in the  $s^2$  term and  $c_1$  is in the  $s$  term. One way to determine the zeros of the model then is to make the effect of the zeros on the characteristic equation the same for both the model and the given system. For example, if the given system has two zeros, then  $c_2$  is some proportion of the  $s^2$  term and  $c_1$  is some proportion of the  $s$  term. To specify the zeros of the model, set  $c_2$  and  $c_1$  in the same proportion in the model's characteristic equation as they were in the system's characteristic equation. Unfortunately such a

scheme works only when the order of the model is one greater than the number of zeros in the given system. For systems of higher order it is not at all clear what should be done and some other method must be developed.

Another interesting question is whether some physical meaning can be given to the difference in the orientations of the two  $V$  surfaces when their shapes are matched. Remember that the modeling technique involved the matching of the shapes of the surfaces while ignoring their orientations. The modeling process was said to be a good one when the two surfaces were oriented in the same direction and worse when the angle between them increased. Throughout Examples 3-1 to 3-3 (underdamped third order systems) it was apparent that matching the shapes of the  $V$  surfaces matched the frequency and damping factor of the model and the given system. However the responses of the model and the given system were not matched with respect to phase and the model tended to lead the system more and more as the approximation became worse (Figs. 3-7 to 3-9). It would be interesting if this phase difference could somehow be related to the difference in the orientations of the two  $V$  surfaces. Then the behavior of the system would be known exactly by adjusting the phase of the model's response.

While the general modeling technique of matching  $V$  surfaces for one choice of  $\dot{V}$  is an acceptable method for

determining models, it has one big disadvantage--the process requires that matrices be transformed into their diagonal form. It is not hard to transform 2 by 2 matrices into their diagonal form but this process is much more difficult for 3 by 3 matrices. The work is further complicated when the  $\underline{M}$  matrix cannot be obtained in diagonal form by a proper choice of  $\dot{V}$ . Such problems were encountered in the development of the third order model. It would indeed be fortunate if an alternate method could be found for the third order model as it was for the second order model. This alternate method involves taking the determinant of a matrix--a relatively simple process. This approach was tried on third order systems by matching the determinants for  $\dot{V}_1 = -x_1^2$ ,  $\dot{V}_2 = -x_2^2$ ,  $\dot{V}_3 = -x_3^2$ . As was seen, these choices for  $\dot{V}$  gave conflicting requirements on the parameters of the model. Perhaps different choices for  $\dot{V}$  would give nonconflicting conditions accurately determining the model. At present these choices are not apparent.

While this work was approached from the point of view of analysis, the techniques developed can also be applied to the synthesis or compensation of control systems. In analysis the system is completely fixed and the model is completely free. The alternate problem of synthesis has the system free and the model fixed. Usually the system cannot be completely free and only one or more parameters are adjustable. The problem now is given a desired model,

adjust the free parameters of the system to give a  $V$  surface which most closely matches the  $V$  surface of the model. In this context the model is actually the desired system. Such an application could employ a search technique using a digital computer to match the  $V$  surfaces.

An alternate and perhaps more useful approach to the problem of synthesis is the utilization of compensating networks. First the model is obtained for a given fixed control system. Then instead of directly compensating the given system, a network is used to compensate the model. Hopefully, this compensating network has the same desired effect when it is also used with the given system.

Last, because the modeling method is independent of the frequency domain, this technique also suggests itself for use in analyzing nonlinear systems. With nonlinear systems however, the  $V$  surfaces do not necessarily form a nested set but may vary in shape in different regions of the state space. Thus the linear models to this nonlinear system are valid only in regions of the state space containing  $V$  surfaces of similar shape. In another region of operation a different linear model is needed, as expected. In addition, it does not seem unreasonable to attempt to find nonlinear models for nonlinear systems.

APPENDIX A

TABULATED  $\underline{P}$  MATRICES FOR GIVEN  $\underline{Q}$  MATRICES<sup>3</sup>

The following tables give  $\underline{P}_i$  ( $V = \underline{x}^T \underline{P}_i \underline{x}$ ) for  $\underline{Q}_i$  ( $\dot{V} = -\underline{x}^T \underline{Q}_i \underline{x}$ ).  $\underline{Q}_i$  is chosen with  $q_{ii} = 1$  and all other elements equal to zero. The system matrix is assumed to be given in terms of phase variables.

$$\underline{A} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ -a_1 & -a_2 & -a_3 & \dots & -a_n \end{bmatrix}$$

Part I. (Second Order System)

$$\underline{P}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \underline{P}_{-1} = \frac{1}{2a_1 a_2} \begin{bmatrix} (a_1 + a_2^2) & a_2 \\ a_2 & 1 \end{bmatrix}$$

$$\underline{P}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \underline{P}_{-2} = \frac{1}{2a_2} \begin{bmatrix} a_1 & 0 \\ 0 & 1 \end{bmatrix}$$

Part II. (Third Order System)

$$Q_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad P_1 = \frac{1}{2a_1(a_3a_2-a_1)}$$

$$\begin{bmatrix} (a_3a_2-a_2a_1+a_3^2a_1) & a_3^2a_2 & (a_3a_2-a_1) \\ a_3^2a_2 & (a_3^3+a_1) & a_3^2 \\ (a_3a_2-a_1) & a_3^2 & a_3 \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad P_2 = \frac{1}{2(a_3a_2-a_1)}$$

$$\begin{bmatrix} a_3a_1 & a_1 & 0 \\ a_1 & (a_3^2+a_2) & a_3 \\ 0 & a_3 & 1 \end{bmatrix}$$

$$Q_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P_3 = \frac{1}{2(a_3a_2-a_1)}$$

$$\begin{bmatrix} a_1^2 & a_2a_1 & 0 \\ a_2a_1 & (a_3a_1+a_2^2) & a_1 \\ 0 & a_1 & a_2 \end{bmatrix}$$



Part III. (Fourth Order System)

$$R_2 = (a_4 a_3 - a_2); R_3 = (a_4 a_3 a_2 - a_2^2 - a_4^2 a_1)$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\underline{Q}_1 = \frac{P}{1} = \frac{2a_4 R_3}{2a_4 R_3}$$

$$\begin{bmatrix} (a_2 R_3 + a_1 (a_3 R_2 - a_4 a_1)) & a_2 (a_3 R_2 - a_4 a_1) \\ a_2 (a_3 R_2 - a_4 a_1) & ((a_3^2 - a_1) R_2 + a_4 a_1 (a_4^2 - a_3)) \\ (a_4 R_3 + a_1 R_2) & a_4 a_3 R_2 \\ & (a_3 R_2 - a_4 a_1) \end{bmatrix}$$

$$\begin{bmatrix} (a_4 R_3 + a_1 R_2) & R_3 \\ (a_4 a_3 R_2) & (a_3 R_2 - a_4 a_1) \\ (a_4^2 R_2 + a_4 a_1) & a_4 R_2 \\ a_4 R_2 & R_2 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\underline{Q}_2 = \frac{P}{2} = \frac{1}{2} \frac{R_3}{R_3}$$

$$\begin{bmatrix} a_1 R_2 & a_4^2 a_1 & 0 \\ a_4^2 a_1 & (a_3 R_2 - a_4 a_1 + a_4^2 a_2) & R_2 \\ a_4 a_1 & a_4^2 a_3 & (a_4^2 + a_2) a_4 \\ 0 & R_2 & a_4^2 a_4 & a_4 \end{bmatrix}$$

$$\frac{Q_3}{P_3} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \frac{P_3}{2R_3} = \frac{1}{2R_3}$$

$$\begin{bmatrix} a_4 a_1^2 & a_4 a_2 a_1 & a_2 a_1 & 0 \\ a_4 a_2 a_1 & (a_4 a_2 + a_1 R_2) & (a_2^2 + a_4 a_1) & a_4 a_1 \\ a_2 a_1 & (a_2^2 + a_4 a_1) & (a_4 a_2 + a_3 a_2 - a_4 a_1) & a_4 a_2 \\ 0 & a_4 a_1 & a_4 a_2 & a_2 \end{bmatrix}$$

$$\frac{Q_4}{P_4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \frac{P_4}{2R_4} = \frac{1}{2R_4}$$

$$\begin{bmatrix} a_2 a_1^2 & a_1 a_2^2 & a_2 a_1 & 0 \\ a_1 a_2^2 & (a_4 a_1^2 + a_2^3) & a_3 a_2 & a_2 a_1 \\ (a_3 a_2 a_1 - a_4 a_1^2) & a_3 a_2 & a_2 a_1 & 0 \\ 0 & a_2 a_1 & a_2 a_1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} (a_3 a_2 a_1 - a_4 a_1^2) & 0 \\ a_3 a_2^2 & a_2 a_1 \\ (a_4 a_2^2 + a_3 a_2 - a_4 a_3 a_1 - a_2 a_1) & a_2^2 \\ a_2^2 & (a_3 a_2 - a_4 a_1) \end{bmatrix}$$

## APPENDIX B

### FREQUENCY DOMAIN MODELS

It is possible to obtain a second order model of any order system by matching certain frequency domain characteristics of the model and the system. One such method matches the phase margin and the frequency at crossover of the two systems. The resulting model is called the phase margin model and it is used quite extensively in the analysis and synthesis of systems. Usually phase margin techniques are not associated with a model even though the existence of such a model may be indirectly assumed. For example, in designing or analyzing control systems phase margin is used as an indication of the behavior of these systems. From experience with second order systems, a phase margin of some particular value indicates that the higher order system will have a certain response. In other words it is assumed that the higher order system will behave similarly (have the same overshoot) as the second order system which has the same phase margin. This second order system therefore can be thought of as a model of the given control system. When the crossover frequencies are also matched the model and the given system have similar rise times. The second order

system which matches both the phase margin and crossover frequency of the given system is termed the phase margin model. The procedure to obtain the phase margin model is to find the phase margin and crossover frequency of the given control system and then adjust the two variable parameters of the model to yield these same values. An example illustrating this procedure is contained in Example 3-1.

Another method of modeling is based on the root locus technique. The model consists of the dominant closed loop poles of the given system and the remaining poles are ignored. A second order model, for example, consists of only the two most dominant poles. To obtain the model it is necessary to find all of the closed loop poles of the system--a difficult task when the order of the system is high and some poles are complex conjugates.

## LIST OF REFERENCES

1. Bower, J. L. and P. M. Schultheiss, Introduction to the Design of Servomechanisms. New York: John Wiley & Sons, Inc., 1958.
2. Higgins, W. T., "Linear System Optimization by the Second Method of Liapunov," Tucson, Arizona: University of Arizona M.S. Thesis, 1964.
3. Ingwerson, D. R., "A Modified Lyapunov Method for Nonlinear Stability Analysis," Palo Alto: Stanford University Ph.D. Thesis, 1960.
4. Kalman, R. E. and J. E. Bertram, "Control System Analysis and Design Via the 'Second Method' of Liapunov, Part I, Continuous-Time Systems," Trans. of ASME, J. of Basic Engr. (June 1960), 371-393.
5. LaSalle, J. and S. Lefschetz, Stability by Liapunov's Direct Method, New York: Academic Press, 1961.
6. Rekasius, Z. V., "A general Performance Index for Analytical Design of Control Systems," IRE Transactions on Automatic Control, May 1961.
7. Vogt, W. G., "Transient Response from the Lyapunov Stability Equation," 1965 Joint Automatic Control Conference.

242  
64

N66-19885

SECTION III

Linear Systems Design Using  
State Variable Feedbacks

TABLE OF CONTENTS

Chapter		Page
	Acknowledgment . . . . .	iii
	List of Figures . . . . .	v
I	Introduction . . . . .	1
II	System Description and Design Objectives . . . . .	5
III	The Regulator Problem . . . . .	11
IV	The Servomechanism Problem . . . . .	44
V	Conclusions . . . . .	56
	References . . . . .	60

## LIST OF FIGURES

Number		Page
1	Pictorial Interpretation of the Regulator Problem . . . . .	13
2	Compensation Using State Variable Feedback . . . . .	19
3	Block Diagram for Example 3 . . . . .	25
4	An Aid in Understanding Method III . . . . .	28
5	Bode Diagrams for Finding the Feedback Coefficients of Example 4 . . . . .	32
6	Bode Diagrams for Finding the Feedback Coefficients for Example 5 . . . . .	36
7	The Servomechanism Problem . . . . .	45
8	Root Locus Plot for Example 5 . . . . .	51
9	The Compensated Systems of Example 5 . . . . .	53



## ABSTRACT

Modern, time-domain methods are used to discuss the control of linear, constant-coefficient systems with unconstrained control effort. Two rather general Performance Indices are used to define two related problems, the Regulator Problem and the Servomechanism Problem.

The Regulator Problem uses the Performance Index

$$J_r = \int_0^{\infty} \underline{x}' Q \underline{x} dt.$$

The solution to the Regulator Problem requires a control structure which contains an inner loop for each of the state variables in the problem formulation. It is shown that this structure permits control over the closed-loop poles of the optimal system and cancellation of unwanted zeroes, but no new zeroes can be added. Three methods are given for computing explicit optimal control systems for specific examples, the parameter optimization problem is reviewed and discussed in relation to the Regulator Problem, and a method is given for introducing zeroes into the Regulator Problem.

The Servomechanism Problem uses the Performance Index

$$J_s = \int_0^{\infty} \left[ (r - \underline{c}' \underline{x})^2 + u^2 \right] dt.$$

The solution to this problem consists of two parts, the prefilter and the regulator. The prefilter shapes the reference input signal before that signal is applied to the regulator portion of the optimal system. The regulator is found by solving the Regulator Problem.

# CHAPTER I

## INTRODUCTION

The most familiar problem in automatic control is the control of linear, constant-coefficient systems with unconstrained control effort. The conventional approach to this problem utilizes Laplace transform techniques to convert the differential equations of the system into algebraic equations. The resulting transformed equations are often displayed pictorially in block diagram form, and the most widely studied block diagram is the single-loop, unity-ratio configuration. Methods based upon use of the Laplace transform are generally referred to as frequency-domain techniques.

This thesis uses modern, time-domain methods to discuss the familiar linear control problem mentioned above. The time-domain approach requires a system description in terms of first-order differential equations obtained directly from the differential equations describing the system. In keeping with conventional methods the results of a time-domain design may also

ultimately be described in terms of a block diagram. The design results in a configuration differing markedly from that resulting from the use of frequency-domain techniques.

Design specifications for use with frequency-domain methods of design are quite diverse; e.g., bandwidth, per cent overshoot in response to a step input, and velocity error constant. In any given problem, specifications may be given in both the time domain and the frequency domain. In the time-domain approach to system design all performance requirements must be embodied in a single specification called the Performance Index. Two rather general integral Performance Indices are used in this thesis to define two related problems, the Regulator Problem and the Servomechanism Problem.

The Regulator Problem uses the Performance Index

$$J_R = \int_0^{\infty} \underline{x}' Q \underline{x} dt \quad (1-1)$$

The solution to the Regulator Problem requires that the uncompensated system be given an input which is a linear combination of the state variables of the system. This result specifies the structure of the optimal control

system as one with an inner loop for each of the state variables in the problem formulation. Utilizing this structure, it is possible to show that the use of state variable feedback allows the designer to control the locations of the poles of the system and to cancel unwanted zeroes, but no new zeroes can be added. Three methods are given for computing explicit optimal control systems for specific examples, and the limitations of each method are discussed. The parameter optimization is defined and the relationship between this problem and the Regulator Problem is explained. An attempt is made to relate the Regulator Problem to conventional frequency-domain design, by showing how zeroes can be introduced into the Regulator Problem.

The Servomechanism Problem uses the Performance Index

$$J_s = \int_0^{\infty} \left[ (r - \underline{c}'\underline{x})^2 + u^2 \right] dt \quad (1-2)$$

The solution to the Servomechanism Problem is closely related to that of the Regulator Problem. It consists of two parts, the prefilter and the regulator. The task of the prefilter is to shape the reference input signal

before the signal is applied to the regulator portion of the optimal system. The regulator is found by solving the Regulator Problem, and as a result of this obvious connection between the two problems, results obtained for either problem apply, in part, to the remaining one.

Throughout the thesis, a variety of techniques for obtaining explicit numerical solutions is presented. In the interests of both clarity and brevity, full use is made of examples.

CHAPTER II  
SYSTEM DESCRIPTION AND DESIGN OBJECTIVES

In this chapter notation and system representation are discussed, controllability and observability are defined, and the design of linear systems is cast into the framework of the modern state variable approach to optimal control theory.

The physical systems considered here are those which can be adequately characterized by a set of ordinary linear differential equations with constant coefficients. Systems having no input, or forcing function, (autonomous systems) as well as those having a scalar input (non-autonomous systems) are studied. It is assumed that these differential equations have already been written.

In the state variable approach to the design of linear systems, the differential equations of the systems are replaced by a set of first-order differential equations of the form

$$\begin{aligned}\dot{\underline{x}} &= \underline{A}\underline{x} + \underline{b}u \\ \underline{y} &= \underline{C}'\underline{x}\end{aligned}\tag{2-1}$$

where  $\underline{x}$  is an  $n$ -vector, the state of the system

$A$  is an  $n$  by  $n$  matrix of constants, the system matrix

$\underline{b}$  is an  $n$ -vector of constants, the control vector

$u$  is a scalar, the control function

$\underline{y}$  is a  $p$ -vector, the output of the system

$C$  is an  $n$  by  $p$  matrix of constants

Frequently the output  $\underline{y}$  will be a scalar instead of a vector; in this case, the matrix  $C$  is replaced by the vector  $\underline{c}$  and  $y$  becomes a linear combination of the state variables,

$$y = \underline{c}'\underline{x} \quad (2-2)$$

The notation indicated above is used throughout. Underlined lower-case letters refer to vectors, and the elements of the vector are denoted by single subscripts. Lower-case letters that are not underlined indicate scalars or constants. Upper-case letters refer to matrices, and the elements of the matrix are denoted by the corresponding lower-case letter with double subscripts. If  $F$  is an arbitrary matrix, then  $F'$  is its transpose; if  $F$  is square and nonsingular  $F^{-1}$  denotes its inverse. The square matrix  $F$  is called symmetric if  $F' = F$ , and positive definite [positive semidefinite] if  $\underline{x}'F\underline{x}$  is

a positive definite [positive semidefinite] function of  $\underline{x}$ ; i.e., one which is always positive [non-negative] except at  $\underline{x} = 0$ , where it is zero.

The concepts of "controllability" and "observability" introduced by Kalman (1963) are needed for a reasonably general discussion of linear systems. Although of fundamental importance from a mathematical viewpoint, these concepts are sufficiently general that they are usually not a major concern for physical systems. Thus it is sufficient to give brief definitions and explicit criteria for determining whether the system of equations (2-1) represents a completely controllable and completely observable system.

A system is completely controllable if all state variables in the representation (2-1) can be affected by some suitable choice of the control function  $u(t)$ . An equivalent mathematical statement is

$$\text{rank } (\underline{b}, \underline{A}\underline{b}, \dots, \underline{A}^{n-1}\underline{b}) = n \quad (2-3)$$

The expression in parentheses in (2-3) is an  $n$  by  $n$  matrix whose columns are the vectors  $\underline{b}, \underline{A}\underline{b}, \dots, \underline{A}^{n-1}\underline{b}$ . A system is completely observable if all the state variables of the system contribute to the output of the system during



a finite time interval. An equivalent mathematical statement is

$$\text{rank } (C, A'C, \dots, (A')^{n-1}C) = n \quad (2-4)$$

If the matrix  $C$  in (2-1) is replaced by the vector  $\underline{c}$ , then the resulting system is completely controllable and completely observable if and only if the numerator and the denominator of the transfer function  $\underline{c}'(sI - A)^{-1}\underline{b}$  have no common cancellable factors (Leake 1964, p. 10). The transfer function given above is merely the overall transfer function  $y(s)/u(s)$ . This can be verified as follows. Taking the Laplace transform of (2-1) under the assumption of zero initial conditions gives

$$\begin{aligned} s\underline{x}(s) &= A\underline{x}(s) + \underline{b}u(s) \\ y(s) &= \underline{c}'\underline{x}(s) \end{aligned} \quad (2-5)$$

Solving the first equation of (2-5) gives

$$\underline{x}(s) = (sI - A)^{-1}\underline{b}u(s) \quad (2-6)$$

Substituting (2-6) into the equation for  $y(s)$  and forming the ratio  $y(s)/u(s)$  gives the desired result; namely,

$$y(s)/u(s) = \underline{c}'(sI - A)^{-1}\underline{b}$$

In order to put the design of control systems on an analytical basis, a criterion of performance or

Performance Index is introduced. The Performance Index is usually an integral selected by the designer as the best single means of judging the behavior of the system. Once the selection of a Performance Index has been made, the problem is converted to one of applied mathematics, with the object being to minimize the value of the chosen integral.

For example, a possible choice for a Performance Index is the familiar integral of the squared error,

$$J = \int_0^{\infty} [c(t) - r(t)]^2 dt = \int_0^{\infty} e(t)^2 dt \quad (2-7)$$

By integrating the square of the difference between the desired output  $r(t)$  and the system output  $c(t)$ , the Performance Index attempts to characterize the accuracy of the control system. The best system is the one which causes the Performance Index to be minimized; referring to (2-7), the best system is the one whose output is as nearly equal to the desired output as design freedom permits.

The two basic problems considered in this thesis are those of finding control functions  $u$  which give the minimum values of one of two particular types of

Performance Indices. The designer is presented with the set of equations (2-1) and is asked to find the control function  $u$  that minimizes one of the following two Performance Indices:

$$J_R = \int_0^{\infty} (\underline{x}' Q \underline{x} + u^2) dt \quad (2-8)$$

$$J_S = \int_0^{\infty} [(r - \underline{c}' \underline{x})^2 + u^2] dt \quad (2-9)$$

These two problems, known as the Regulator Problem and the Servomechanism Problem are defined more precisely and considered in detail in the ensuing chapters.

CHAPTER III  
THE REGULATOR PROBLEM

In this chapter the first of the two problems introduced at the close of the previous chapter is defined, its general solution is studied in terms of the structure of the optimal system, and three methods are presented for obtaining specific solutions to the design problem. The relationship between the Regulator Problem and the parameter optimization problem is discussed and a method for introducing zeroes is described.

The following statement of the Regulator Problem is adapted from that of R. J. Leake (1964, pp. 4-5):

Regulator Problem For the system (2-1) where  $\underline{x}(0)$  represents a set of nonzero values of the state variables at  $t = 0$ , find a continuous control function  $u$  such that the system is transferred from its initial state  $\underline{x}(0)$  to the origin of the state space in such a way that the Performance Index

$$J_r = \int_0^{\infty} (\underline{x}' Q \underline{x} + u^2) dt \quad (3-1)$$

is minimized. The matrix  $Q$  is symmetric and positive definite or semidefinite.

A pictorial interpretation of the Regulator Problem is shown in Fig. 1. The figure shows a block labelled "controller" having as its inputs the state variables and an as yet unspecified vector function  $\underline{r}(t)$ . These two vectors are combined to produce the scalar control function  $u$  which transfers the state of the system to the origin while minimizing (3-1).

The central theorem relating to the Regulator Problem, proved in Kalman (1964), is presented below.

Theorem I Assume that (2-1) represents the equations of motion of a completely controllable system and define a Performance Index

$$J(T) = \int_0^T (\underline{x}' Q \underline{x} + u^2) dt \quad (3-2)$$

where the matrix  $Q$  is symmetric and positive definite or semidefinite. Let  $P(t) = \mathbb{T}(t; T, 0)$  be the unique  $n$  by  $n$  symmetric matrix solution (where the parameters  $T$  and  $0$  correspond to the upper and lower limits in (3-2)) of the matrix Riccati differential equation

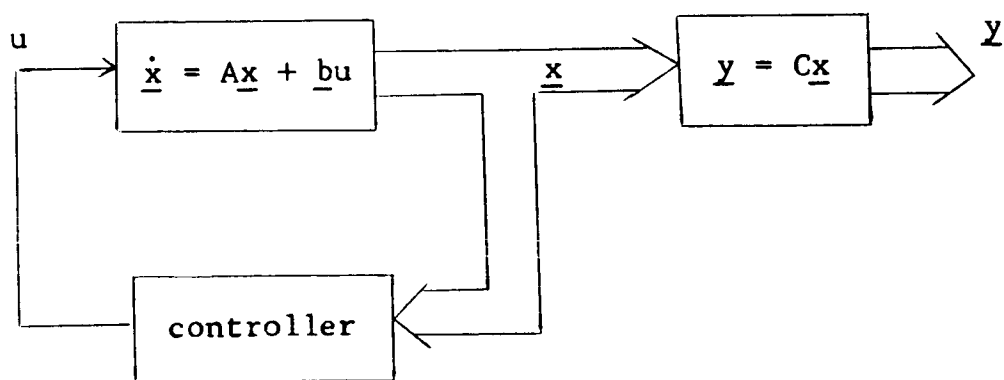


Fig. 1 Pictorial Interpretation of the Regulator Problem

$$-\frac{dP}{dt} = PA + A'P - P\underline{b}\underline{b}'P + Q \quad (3-3)$$

satisfying the boundary conditions

$$P(T) = \Pi(T;T,0) = 0. \quad (3-4)$$

Then the optimal control function for the Regulator Problem exists, is unique, and is given by

$$u^0 = -\underline{x}'P_0\underline{b} = -\underline{k}'\underline{x} \quad (3-5)$$

where

$$P_0 = \lim_{T \rightarrow \infty} \Pi(0;T,0) \quad (3-6)$$

and

$$\underline{k} = P_0\underline{b} \quad (3-7)$$

In addition, the minimum value of  $J_r$  is given by

$$J_r^0 = \underline{x}'(0)P_0\underline{x}(0) \quad (3-8)$$

Furthermore, if the system is completely observable as well as completely controllable and if the matrix  $Q$  in (3-1) and (3-2) is replaced by  $CC'$ ,  $P_0$  is positive definite and the resulting optimal system is asymptotically stable.

Theorem I states that complete controllability is a sufficient condition for the existence of a solution to design problems which use Performance Indices of the form (3-1). For systems which are also completely

observable, the theorem assures the designer that he will wind up with a compensated system which is stable.

The theorem also shows why an output equation ( $\underline{y} = C'\underline{x}$ ) is included in the system equations. The output equation arises as a result of writing the matrix  $Q$  as

$$Q = CC' \quad (3-9)$$

In the usual design procedure the designer decides on a positive definite or semidefinite matrix  $Q$ ; and as an aid to his intuition the resulting quadratic form can be written as

$$\underline{x}'Q\underline{x} = (\underline{x}'C) (C'\underline{x}) = \underline{y}'\underline{y} = \|\underline{y}\|^2 \quad (3-10)$$

so that  $J_R$  becomes

$$J_R = \int_0^{\infty} (\|\underline{y}\|^2 + u^2) dt \quad (3-11)$$

The equation (3-4), where the elements of  $\underline{k}$  are referred to as the "feedback coefficients," expresses the fact that the optimum control function for the Regulator Problem is a linear combination of the state variables. This is a highly important result, as it specifies the form of the optimum system. To illustrate this consider the following example:

#### Example 1

It is desired to compensate a system having the following uncompensated overall transfer function



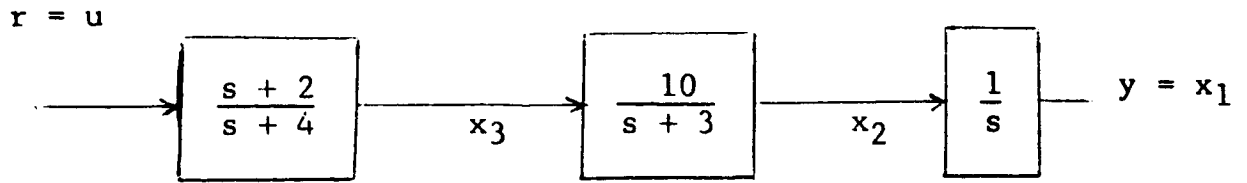
$$\frac{y(s)}{u(s)} = \frac{10(s + 2)}{(s + 4)(s + 3)(s)}$$

where  $y(s)$  is the Laplace transformed output and  $u(s)$  is the control function. A block diagram for the uncompensated system is shown in Fig. 2(a). Note that the system is non-autonomous, since it has an input  $u$ . If the design were carried out by using a Performance Index of the form (3-1) and the state variables shown on the figure, then the compensated system would have the appearance of Fig. 2(b). In Fig. 2(b) an input  $r$  has been added to keep the system non-autonomous.

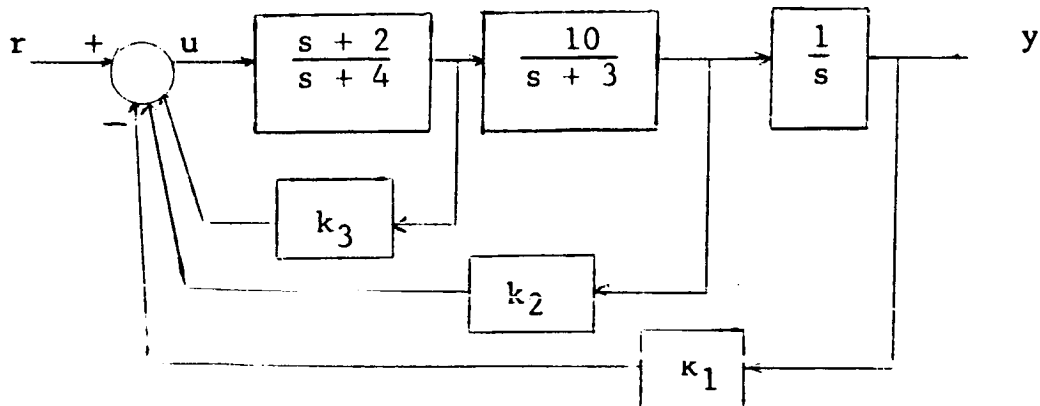
The numerical values of the feedback coefficients for Example 1 are as yet unspecified since no procedures for calculating the  $k_i$  (feedback coefficients) have yet been given. However, the overall transfer function for the compensated system is easily seen to be

$$\frac{y(s)}{r(s)} = \frac{10(s + 2)}{(k_3 + 1)s^3 + (5k_3 + 10k_2 + 7)s^2 + (6k_3 + 20k_2 + 10k_1 + 12)s + 20k_1}$$

so that by proper choices for  $k_1$ ,  $k_2$ , and  $k_3$  any three desired pole locations can be obtained, but the zeroes remain unchanged. This complete control of the pole locations and lack of control of the zero locations is



a) Uncompensated System



b) Compensated System

Fig. 2 Compensation Using State Variable Feedback

a general consequence of compensation by state variable feedback (Brockett 1965).

That the first part of the previous statement is true may be seen by considering a more general system with the overall transfer function

$$\frac{y(s)}{u(s)} = \frac{c_n s^{n-1} + c_{n-1} s^{n-2} + \dots + c_1}{s^n + a_n s^{n-1} + \dots + a_1} \quad (3-12)$$

Representing the system in phase variables the equations of motion become

$$\begin{aligned} \underline{x} &= \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & \dots & \\ 0 & 0 & 0 & \dots & 1 \\ -a_1 & -a_2 & -a_3 & \dots & -a_n \end{bmatrix} \underline{x} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u \\ y &= \begin{bmatrix} c_1 & c_2 & c_3 & \dots & c_n \end{bmatrix} \underline{x} \end{aligned} \quad (3-13)$$

The assumptions of controllability and observability assure that any single input, single output system has this unique phase variable representation; see Kalman (1963).

Suppose that the characteristic equation associated with the desired pole configuration is  $s^n + r_n s^{n-1} + \dots + r_1$ .

If the control function  $u$  is set equal to  $-\underline{k}'\underline{x}$  with the  $k_i$  defined by

$$k_1 = r_1 - a_1, k_2 = r_2 - a_2, \dots, k_n = r_n - a_n \quad (3-14)$$

then in the autonomous system each  $-a_i$  in the system matrix will be replaced by  $-a_i - (r_i - a_i)$  or  $-r_i$ , and the overall transfer function will have the desired pole configuration. Thus the use of state variable feedback gives the designer both complete control over the pole locations of the system and the means by which unwanted zeroes can be cancelled.

In the above system representation it is clear that no new zeroes can be added by using state variable feedback. Brockett (1965) shows that this is always the case; namely, for a completely controllable and observable system using state variable feedback, no new zeroes can be added.

In one sense it is unfortunate that no zeroes can be added since the most common forms of linear compensation (lead, lag, and lead-lag) require at least one zero in the compensator. In another sense, however, zeroes are not necessary because for a given Performance Index of the form (2-6) Theorem I guarantees that the designer can always get the best design by feeding back all the state variables. However, if a zero is included in the

forward transfer function, as with a lead or a lag network, then the pole of the lead or lag compensator will cause an increase in the order of the system. It is entirely possible, indeed probable, that the value of the Performance Index for the optimized  $n + 1$ -order system may be less than that for the optimal  $n$ th order system. A means for introducing zeroes into the Regulator Problem is given later on in this chapter.

There are three methods for finding the elements of  $\underline{k}$ , the coefficients of the state variables in the expression for the optimal control function in the Regulator Problem.

Method I      Solution of the Ricatti Equation

Method II     Kalman's Equation

Method III    Bode Diagram Design

#### Method I    Solution of the Ricatti Equation

The first method for calculating the feedback coefficients requires the solution of (3-3) in Theorem I for the unknown symmetric matrix  $P$ . Once  $P$  is known,  $P_0$  is found by calculating the limit in (3-6) and then the feedback coefficients are obtained from (3-7). The difficult step in this procedure for finding  $\underline{k}$  is in

obtaining a solution to the nonlinear Riccati differential equation. Even for second-order systems the matrix differential equation is difficult to solve (Leake 1964), and so numerical techniques must be used, although numerical solutions are not shown here.

An algebraic matrix equation which can be used to find  $P_0$  is also available. By (3-6)  $P_0$  is an equilibrium state of the Riccati equation; accordingly, by setting

$$\frac{dP}{dt} = 0 \text{ in (3-3)}$$

$$P_0 A + A' P_0 - P_0 \underline{bb}' P_0 + Q = 0 \quad (3-15)$$

Hand solution of (3-15) is tractable for second-order and even third-order systems. The equation is difficult to solve by numerical techniques because it is a set of nonlinear equations and thus the solutions are not unique. Kalman (1964) shows that (3-15) has a unique solution which is identical to the solution of the Riccati differential equation if the system is completely observable and  $P_0$  is positive definite.

### Example 2

Consider the design of the first-order system

$$\dot{x} = -x + u$$

through the use of the Performance Index

$$J = \int_0^{\infty} (x^2 + u^2) dt.$$

The Ricatti differential equation is

$$-\frac{dP}{dt} = PA + A'P - P\underline{b}b'P + Q$$

which in this first-order case is the scalar equation

$$\begin{aligned} -\frac{dp}{dt} &= (-1)p + (-1)p - p(1)(1)p + 1 \\ &= -2p - p^2 + 1 \end{aligned}$$

along with the boundary condition

$$p(T) = \pi(T; T, 0) = 0$$

Using the technique described by Leake (1964) the

analytical solution to the above equation is

$$p(t) = \frac{\exp(-\sqrt{2}(t-T)) - \exp(\sqrt{2}(t-T))}{(\sqrt{2} + 1)\exp(\sqrt{2}(t-T)) + (\sqrt{2} - 1)\exp(\sqrt{2}(t-T))}$$

Taking the limit prescribed in (3-6)

$$\begin{aligned} p_0 &= \lim_{T \rightarrow \infty} \frac{\exp(\sqrt{2}T) - \exp(-\sqrt{2}T)}{(\sqrt{2} + 1)\exp(\sqrt{2}T) + (\sqrt{2} - 1)\exp(-\sqrt{2}T)} \\ &= \sqrt{2} - 1 \end{aligned}$$

From (3-7)

$$\begin{aligned} k &= p_0 b \\ &= \sqrt{2} - 1 \end{aligned}$$

Using the algebraic Ricatti equation for this example leads to a quick solution. The equation (3-15) becomes

$$-2p_0 - p_0^2 + 1 = 0$$

and the unique positive definite solution is

$$p_0 = \sqrt{2} - 1$$

### Method II Kalman's Equation

Kalman (1964) has found an algebraic equation which can be solved directly for the feedback coefficients, without first finding the matrix  $P_0$ . This equation offers some computational advantages over the algebraic Ricatti equation; moreover, it is a frequency-domain equation and thus provides a link between conventional and modern control theory.

The derivation of Kalman's Equation starts with the algebraic Ricatti Equation (3-15), rewritten as

$$-P_0 A - A' P_0 = CC' - P_0 \underline{b} \underline{b}' P_0. \quad (3-16)$$

Adding and subtracting  $sP_0$  gives

$$P_0(sI - A) + (-sI - A')P_0 = CC' - P_0 \underline{b} \underline{b}' P_0. \quad (3-17)$$

Letting

$$\phi(s) = (sI - A)^{-1} \quad (3-18)$$

and multiplying (3-17) from the left by  $\underline{b}'\phi'(-s)$  and

from the right by  $\phi(s)\underline{b}$  gives

$$\underline{b}'\phi'(-s)P_0\underline{b} - \underline{b}'\phi(s)P_0\underline{b} = \underline{b}'\phi'(-s) [CC' - P_0\underline{b}\underline{b}'P_0] \phi(s)\underline{b}. \quad (3-19)$$



From Theorem I,

$$\underline{k} = P_0 \underline{b} \quad (3-20)$$

so that (3-19) becomes

$$\underline{b}' \phi'(-s) \underline{k} - \underline{b}' \phi(s) \underline{k} = \underline{b}' \phi(-s) C C' \phi(s) \underline{b} - \underline{b}' \phi'(-s) \underline{k} \underline{k}' \phi(s) \underline{b} \quad (3-21)$$

Transposing the second term on the right-hand side of

(3-21) and adding 1 to both sides gives

$$\left[ 1 + \underline{k}' \phi(-s) \underline{b} \right] \left[ 1 + \underline{k}' \phi(s) \underline{b} \right] = 1 + \underline{b}' \phi'(-s) C C' \phi(s) \underline{b} \quad (3-22)$$

or

$$\left| 1 + \underline{k}' \phi(j\omega) \underline{b} \right|^2 = 1 + \left\| C' \phi(j\omega) \underline{b} \right\|^2 \quad (3-23)$$

This is Kalman's Equation. The matrix  $\phi(s)$  is called the resolvent of A and is equal to the Laplace Transform of the state transition matrix.

### Example 3

Compensate the system shown in Fig. 3(a) by using the Performance Index

$$J = \int_0^{\infty} (x_1^2 + u^2) dt.$$

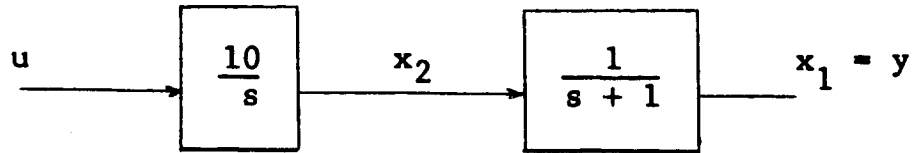
The equations of motion are

$$\dot{\underline{x}} = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \underline{x} + \begin{bmatrix} 0 \\ 10 \end{bmatrix} u$$

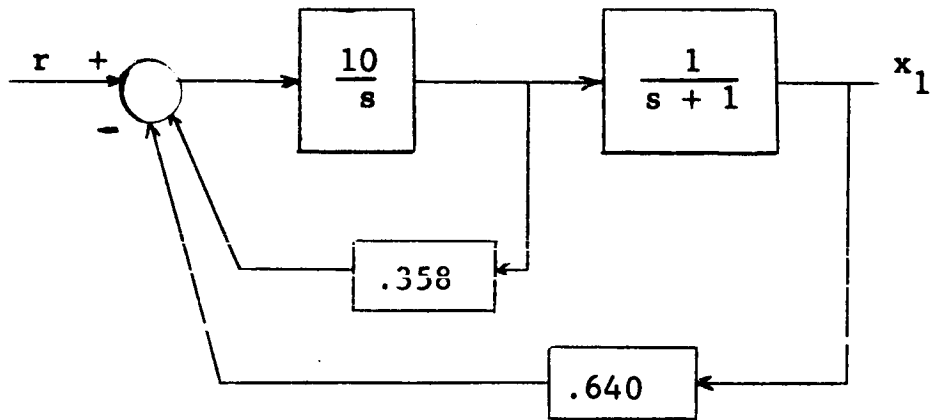
$$c = \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{x}$$

Using (3-18) the resolvent of A is

$$\phi(s) = \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s(s+1)} \\ 0 & \frac{1}{s} \end{bmatrix}$$



a) Uncompensated System



b) Compensated System

Fig. 3 Block Diagram for Example 3

giving

$$\left| 1 + \underline{k}' \phi(j\omega) \underline{b} \right|^2 = \frac{\omega^4 + (100k_2^2 - 20k_1 + 1)\omega^2 + 100k_1^2 + 200k_1k_2 + 100k_2^2}{\omega^4 + \omega^2}$$

and

$$1 + \left| \underline{c}' \phi(j\omega) \underline{b} \right|^2 = \frac{\omega^4 + \omega^2 + 100}{\omega^4 + \omega^2}$$

For (3-23) to hold for all  $\omega$  the following equations must hold:

$$100k_2^2 - 20k_1 = 0$$

$$100k_1^2 + 200k_1k_2 + 100k_2^2 = 100$$

Solving the above two equations yields the numerical values for the feedback coefficients,

$$k_1 = .640, k_2 = .360$$

The compensated system is shown in Fig. 3(b).

### Method III Bode Diagram Design

This third method for calculating the feedback coefficients makes use of a special case of Kalman's Equation; namely, the case in which the matrix  $C$  in (3-23) is replaced by the vector  $\underline{c}$ , so that the Performance Index becomes

$$J = \int_0^{\infty} (\underline{x}' \underline{c} \underline{c}' \underline{x} + u^2) dt \quad (3-24)$$

$$= \int_0^{\infty} [(c_1x_1 + c_2x_2 + \dots + c_nx_n)^2 + u^2] dt \quad (3-25)$$

Besides the assumption that  $J_r$  is of the form given in (3-25) it will also be necessary to assume that the fixed plant contains at least one pure integration.<sup>1</sup>

The block diagram of Fig. 4 is useful for discussing Method III. From the diagram  $u = r_0 - \underline{k}'\underline{x}$ . With this substitution in (2-1) the system equations become

$$\begin{aligned}\dot{\underline{x}} &= \underline{A}\underline{x} + \underline{b}(r_0 - \underline{k}'\underline{x}) \\ &= \underline{A}_k\underline{x} + \underline{b}r_0\end{aligned}\quad (3-26)$$

$$y = \underline{c}'\underline{x}\quad (3-27)$$

where  $\underline{A}_k = \underline{A} - \underline{b}\underline{k}'$ . The input to the block labelled  $G(s)$  is  $u(s)$ , while the output is  $\underline{c}'\underline{x}(s)$ , giving

$$\begin{aligned}G(s) &= \frac{\underline{c}'\underline{x}(s)}{u(s)} \\ &= \frac{\underline{c}'\phi(s)\underline{b}u(s)}{u(s)} \\ &= \underline{c}'\phi(s)\underline{b}\end{aligned}\quad (3-28)$$

Similarly, the transfer function  $H(s)$  is given by

$$H(s) = \frac{\underline{k}'\underline{x}(s)}{\underline{c}'\underline{x}(s)}\quad (3-29)$$

and the negative loop gain,  $A(s)$ , is given by the product of (3-28) and (3-29),

---

1. This is more restrictive than need be; in more precise terms, it is necessary to assume that  $G(s)$  has high gain at low frequencies and low gain at high frequencies.

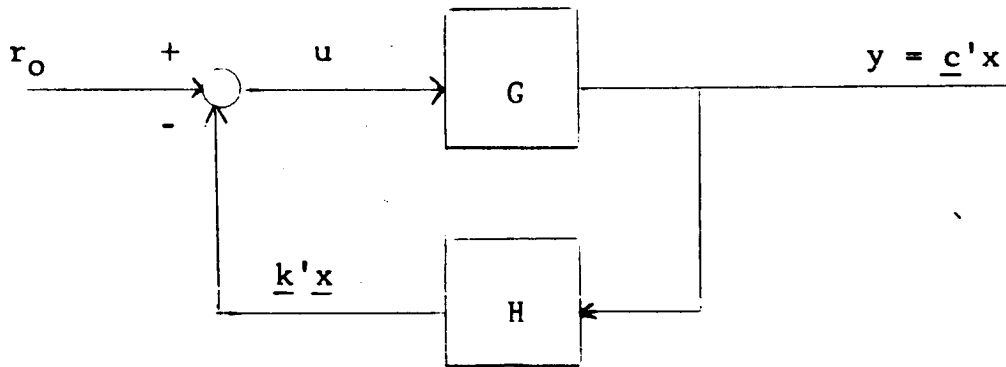


Fig. 4 An Aid in Understanding Method III

$$\begin{aligned}
A(s) &= G(s)H(s) \\
&= \underline{c}'\phi(s)\underline{b} \cdot \frac{\underline{k}'\underline{x}(s)}{\underline{c}'\underline{x}(s)} \\
&= \underline{c}'\phi(s)\underline{b} \cdot \frac{\underline{k}'\phi(s)\underline{b}u(s)}{\underline{c}'\phi(s)\underline{b}u(s)} \\
&= \underline{k}'\phi(s)\underline{b} \tag{3-30}
\end{aligned}$$

The overall transfer function  $y(s)/r_o(s)$ , designated  $M_1(s)$ , is found from (3-26) to be

$$M_1(s) = \underline{c}'\phi_k(s)\underline{b} \tag{3-31}$$

where  $\phi_k(s) = (sI - A_k)^{-1}$ .

Substituting (3-28) and (3-30) into Kalman's Equation (3-23), gives

$$|1 + A(j\omega)|^2 = 1 + |G(j\omega)|^2 \tag{3-32}$$

as the equation which determines the values of the elements of  $\underline{k}$ .

In the design procedure using Method III three properties of (3-32) are used. These properties hold when  $G(s)$  has at least one integration.

- a) for small  $\omega$ ,  $|1 + GH(j\omega)| \doteq G(j\omega)$
- b) for large  $\omega$ ,  $|1 + GH(j\omega)| \doteq 1$
- c) when  $G(j\omega) = 1$ ,  $|1 + GH(j\omega)| = \sqrt{2}$

The design procedure consists of using the Bode diagram of  $G(j\omega)$  and the three properties above to obtain a

good approximation to  $1 + GH(j\omega)$ , and then finding  $\underline{k}$ .

The steps are

- Step 1 Sketch the Bode diagram of  $G(j\omega)$ .
- Step 2 For values of  $\omega$  less than the unity crossover frequency of  $G(j\omega)$ , match the Bode diagram of  $1 + GH(j\omega)$  with that of  $G(j\omega)$ .
- Step 3 For values of  $\omega$  greater than the unity crossover frequency of  $G(j\omega)$ , make the Bode diagram of  $1 + G(j\omega)$  be constant at the value 1; this can be accomplished by using a Butterworth polynomial of the same order as the magnitude of the slope of the Bode diagram of  $G(j\omega)$  at crossover.<sup>2</sup>
- Step 4 By using the results of the previous three steps form an approximate expression for  $1 + GH(s)$  and equate it to the true analytical expression to evaluate  $\underline{k}$ .

This procedure is illustrated in the following example.

---

2. Choosing the characteristic frequency of the Butterworth polynomial as the crossover frequency assures that at crossover the Butterworth polynomial has a magnitude of  $\sqrt{2}$ , so that property (c) will be satisfied.

Example 4

Consider the uncompensated system

$$G(s) = \frac{1438}{s(s^2 + 3.25s + 3.56)}$$

whose phase variable representation is

$$\dot{\underline{x}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -3.56 & -3.2 \end{bmatrix} \underline{x} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u$$

$$y = 1438x_1$$

The Performance Index according to (3-25) is

$$J_r = \int_0^{\infty} [(1438x_1)^2 + u^2] dt$$

To carry out Step 1 write

$$G(s) = \frac{404}{s \left( \frac{s^2}{3.46} + \frac{2(.846)}{1.89}s + 1 \right)}$$

The Bode diagram appears in Fig. 5. Following Step 2 the low frequency part of  $1 + GH(s)$  is given by

$$\frac{1438}{s(s^2 + 3.2s + 3.56)}$$

For frequencies greater than the crossover frequency  $\omega_c$  Step 3 requires that  $|1 + G(j\omega)H(j\omega)|$  be 1; therefore a third-order Butterworth polynomial is chosen, namely

$$\frac{1}{\omega_c^3} (s^3 + 2\omega_c s^2 + 2\omega_c^2 s + \omega_c^3)$$



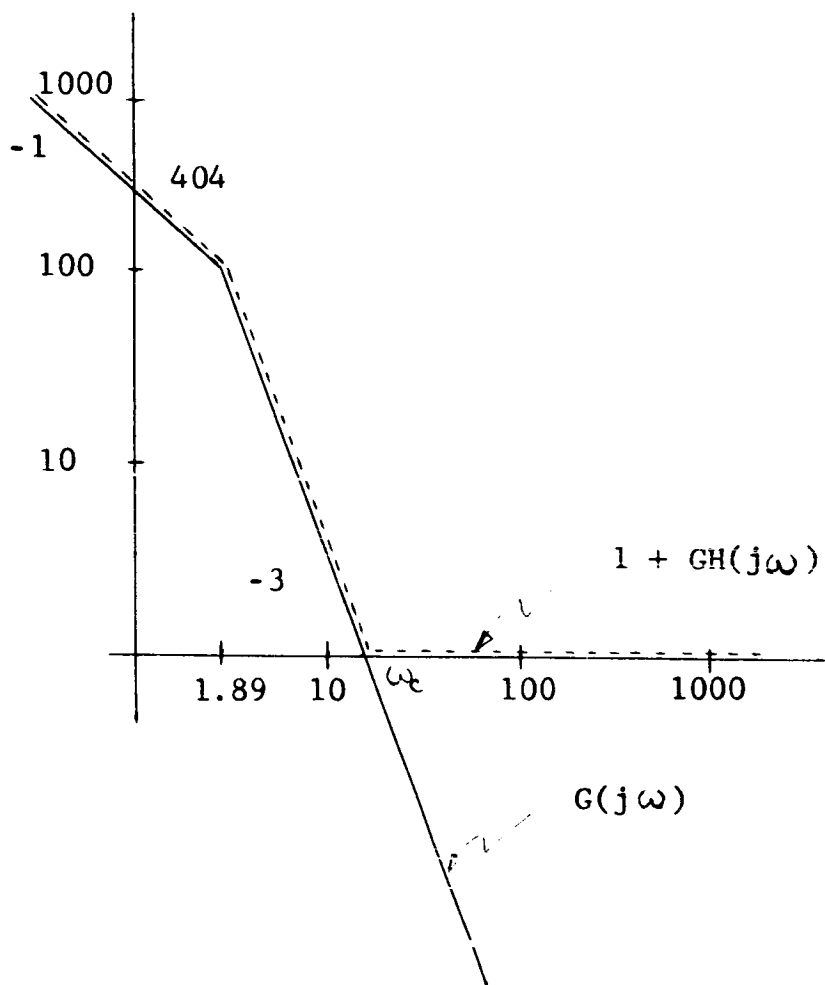


Fig. 5 Bode Diagrams For Finding the Feedback Coefficients of Example 4

The quantity  $1 + G(s)H(s)$  is then approximated by

$$\frac{1438}{s(s^2 + 3.2s + 3.56)} \cdot \frac{1}{\omega_c^3} (s^3 + 2\omega_c s^2 + 2\omega_c^2 s + \omega_c^3) \quad (3-33)$$

It is now clear that  $|1 + G(j\omega)H(j\omega)| = \sqrt{2}$  when  $s = j\omega_c$ , since the first factor in (3-33) has a magnitude of 1 at crossover and the second factor (the Butterworth polynomial) has the magnitude  $\sqrt{2}$ ; thus property (c) is satisfied.

The analytical expression for  $1 + GH(s)$  is found after forming  $H(s)$  through the use of (3-29),

$$\begin{aligned} H(s) &= \frac{k_1 x_1(s) + k_2 x_2(s) + k_3 x_3(s)}{1438 x_1(s)} \\ &= \frac{k_3 s^2 + k_2 s + k_1}{1438} \end{aligned} \quad (3-34)$$

Thus

$$1 + G(s)H(s) = \frac{s^3 + (3.2 + k_2)s^2 + (3.56 + k_2)s + k_1}{s^3 + 3.2s^2 + 3.56s}$$

Step 4 is carried out by finding the values of  $k_1$ ,  $k_2$ ,  $k_3$  and  $\omega_c$  which cause (3-33) and (3-34) to be equal. Those values are

$$\omega_c = \sqrt[3]{1438} = 11.3$$

$$k_1 = 1438, \quad k_2 = 252, \quad k_3 = 19.3$$

By comparison, a digital computer solution of the Riccati differential equation yields  $k_1 = 1437.93$ ,  $k_2 = 251.8$ ,  $k_3 = 19.47$ .

In Example 4 phase variables are used and as a result the designer has no opportunity to choose the Performance Index - he has to accept as the output of the system a linear combination of the derivatives of  $\underline{x}$  in which the weighting factors for the derivatives are determined by the zeroes of  $G(s)$ . A more serious objection to the use of phase variables is that these variables represent successive derivatives of  $x_1$  and for systems whose order exceeds two are not physically available for use as inputs to the linear amplifiers in the inner feedback loops of the compensated system.

Method III can be applied to some systems which are expressed in terms of variables which do not represent successive derivatives. The essential requirement for the successful application of the method is that the approximate expression for  $1 + GH(s)$  have the same form as the true analytical expression. The presence of zeroes in  $G(s)$  makes it impossible to match the approximate expression with the analytical expression, unless phase variables are used. When there are no zeroes present the method will be successful for some choices of the Performance Index; this is shown in the following example.

Example 5

Consider the same system that was used in Example 3, in which phase variables were not used. The uncompensated transfer function is

$$G(s) = \frac{10}{(s)(s + 1)}$$

and using (3-29),  $H(s)$  is found to be

$$H(s) = \frac{k_2 s + k_1 + k_2}{1}$$

so that

$$GH(s) = \frac{10(k_2 s + k_1 + k_2)}{s(s + 1)}$$

The Bode diagram for  $G(j\omega)$  is shown in Fig. 6. From the diagram the low-frequency part of  $1 + GH(s)$  is given by  $G(s)$ , and a second-order Butterworth polynomial is needed for the proper high-frequency behavior. Thus the approximate expression for  $1 + GH(s)$  is

$$1 + GH(s) = \frac{10}{\omega_c^2} \cdot \frac{s^2 + \sqrt{2} \omega_c s + \omega_c^2}{(s)(s + 1)}$$

The analytical expression for  $1 + GH(s)$  is

$$1 + GH(s) = \frac{s^2 + (10k_2 + 1)s + 10(k_1 + k_2)}{(s)(s + 1)}$$

Making both expressions identical requires

$$\omega_c = 3.16, k_1 = .652, k_2 = .348$$

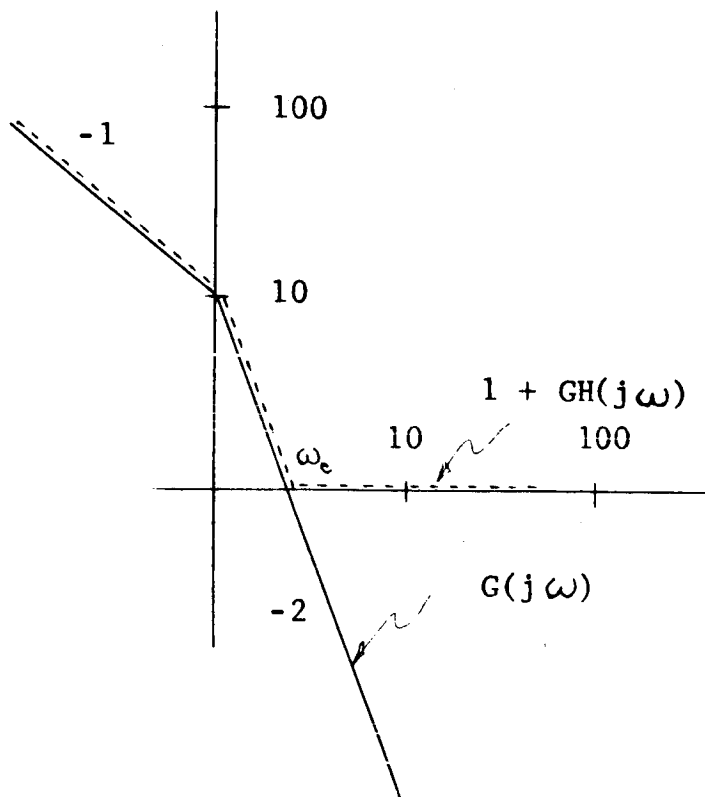


Fig. 6 Bode Diagrams For Finding the Feedback Coefficients of Example 5

These values of the feedback coefficients compare favorably with the more accurate values given in Example 3.

In the preceding example we were able to apply Method III, despite the fact that phase variables were not used. The Performance Index was

$$J = \int_0^{\infty} (x_1^2 + u^2) dt$$

However, if the Performance Index is

$$J = \int_0^{\infty} [(x_1 + x_2)^2 + u^2] dt \quad (3-35)$$

then the method cannot be applied because matching the analytical expression and the approximate expressions for  $1 + GH(s)$  requires

$$\frac{s^3 + 2s^2 + (10k_2 + 1)s + 10k_1}{(s)(s + 1)^2} = \frac{10}{\omega_c^2} \cdot \frac{s^2 + \sqrt{2} \omega_c s + \omega_c^2}{(s)(s + 1)}$$

This is not possible, as the systems are of different order. For a further discussion of Method III see Leake (1965).

The Regulator Problem has been discussed in some detail; next, this design problem will be related to the parameter optimization problem. The reduced Ricatti equation, (3-15), was used in the development of Method II. It is also the starting point for demonstrating the connection between the Regulator Problem and the design of linear systems using quadratic Performance Indices of the form

$$J = \int_0^{\infty} \underline{x}' D \underline{x} dt \quad (3-36)$$

where  $D$  is a positive definite or semidefinite matrix (Kalman 1964). Before the connection is presented, it is necessary to discuss the time-domain procedure for solving the problem associated with (3-36), frequently called the parameter optimization problem.

The procedure for solving the parameter optimization problem consists of two parts: first, the evaluation of (3-36) in terms of the elements of  $D$  and the system matrix of (2-1); second, the selection of those values of the adjustable parameters (as specified in the system matrix) that give the minimum value of the Performance Index.

To carry out the first part the integrand in (3-36) is set equal to a positive definite function of  $\underline{x}$ ,

$$-\dot{V} = \underline{x}' D \underline{x} \quad (3-37)$$

Then the integral becomes

$$J = \int_0^{\infty} -\dot{V}(\underline{x}) dt = -V(\underline{x}) \Big|_{\underline{x}(0)}^{\underline{x}(\infty)} = -V(\underline{x}(\infty)) + V(\underline{x}(0)) \quad (3-38)$$

But for an asymptotically stable system,  $\underline{x}(\infty) = 0$  and the value of  $J$  becomes

$$J = V(\underline{x}(0)) \quad (3-39)$$

$V$  must itself be a positive definite function; that is,  $V$  can be represented by

$$V = \underline{x}' P_0 \underline{x} \quad (3-40)$$

where the matrix  $P_0$  is positive definite and symmetric. Taking the time derivative of (3-40) and comparing the result with the integrand of (3-36) gives

$$A' P_0 + P_0 A = -D \quad (3-41)$$

After solving this set of linear algebraic equations for the elements of the matrix  $P_0$ , (3-39) is used to write the expression for  $J$  in terms of the adjustable parameters and the initial values for the state vector  $\underline{x}$ .

The second part of the design procedure for the parameter optimization problem consists of setting the partial derivatives of  $J$  with respect to the adjustable parameters equal to zero, and solving the resulting nonlinear equations for the optimum values of the adjustable parameters.

Theorem I of this chapter guarantees that the optimal control function for the Regulator Problem is a linear combination of the state variables,

$$u = -\underline{k}' \underline{x} \quad (3-42)$$

where the elements of  $\underline{k}$  are as yet unspecified. If (3-42) is substituted into  $J_r$ , it becomes



$$\begin{aligned}
J_R &= \int_0^{\infty} (\underline{x}'Q\underline{x} + u^2)dt \\
&= \int_0^{\infty} (\underline{x}'Q\underline{x} + \underline{x}'\underline{k}\underline{k}'\underline{x})dt \\
&= \int_0^{\infty} \underline{x}'(Q + \underline{k}\underline{k}')\underline{x}dt \tag{3-43}
\end{aligned}$$

Note that (3-43) and (3-36) are of the same form.

Substituting (3-42) into the system equations (2-1) gives

$$\begin{aligned}
\dot{\underline{x}} &= A\underline{x} + \underline{b}u \\
&= (A - \underline{b}\underline{k}')\underline{x} \tag{3-44}
\end{aligned}$$

so that the feedback coefficients can be considered as adjustable parameters in the parameter optimization problem defined by (3-43) and (3-44). For this parameter optimization problem the equation corresponding to (3-41) is

$$(A - \underline{b}\underline{k}')'P_0 + P_0(A - \underline{b}\underline{k}') = -Q - \underline{k}\underline{k}' \tag{3-45}$$

The above equation can be made the same as the reduced Ricatti equation by the following steps. First, expanding the left-hand side, there results

$$A'P_0 - \underline{k}\underline{b}'P_0 + P_0A - P_0\underline{b}\underline{k}' = -Q - \underline{k}\underline{k}'$$

Utilizing the substitution given in Theorem I,

$$\underline{k} = P_0\underline{b}$$

gives

$$A'P_0 - \underline{k}\underline{k}' + P_0A - P_0\underline{b}\underline{b}'P_0 = -Q - \underline{k}\underline{k}'$$

or

$$P_0A + A'P_0 - P_0\underline{b}\underline{b}'P_0 + Q = 0$$

which is the reduced Ricatti equation. Therefore, considering the Regulator Problem as a parameter optimization problem leads to the same set of equations as given in Theorem I.

There is another link between the Regulator Problem and the parameter optimization problem. In the latter, the system being designed has a fixed form with adjustable gains and time constants. To write the expression for the Performance Index, (3-36), it is necessary to choose a set of initial values for the state vector  $\underline{x}$ . The initial conditions usually chosen are those which make the response of the autonomous system to these initial conditions identical to the error response of that same system to a step input (Gibson 1963).

Now if the fixed form of the system were that of the Regulator Problem - with all state variables being fed back through linear amplifiers - then the values of the adjustable parameters (the feedback coefficients) would be independent of the choice of initial conditions. This suggests that for both the Regulator Problem and the parameter optimization problem the form of the optimum system is the same.

Engineers familiar with conventional, frequency-domain compensation may object to the form of the optimum system discussed above because no new zeroes can be added by using state variable feedback, while even the simplest lead and lag compensators introduce new zeroes. This objection cannot be fully refuted although there is a way of introducing adjustable zeroes into the Regulator Problem.

Adjustable zeroes can be included in the Regulator Problem by putting in tandem with the fixed plant a compensator of the form

$$G_c = \frac{s + z}{s + p} \quad (3-46)$$

In (3-46) the numerator represents the adjustable zero and the denominator a term included to make  $G_c$  realizable. If the uncompensated system is of order  $n$ , the compensated system has order  $n + 1$ . In the compensated system there are  $n + 1$  state variables,  $n + 1$  feedback coefficients, and, in addition, an adjustable zero and an adjustable pole. The best compensated system is the one which yields the lowest value of some chosen Regulator Performance Index; the best compensated system is specified by  $k_1, k_2, \dots, k_n, z$ , and  $p$ .

The computational problems posed by this method of introducing zeroes are formidable, since the addition of the compensator pole and zero causes the plant (the matrix  $A$  in (2-1)) to be incompletely specified. The only feasible way of treating this problem seems to be to use a digital computer to perform a two-dimensional search, looking for the values of  $z$  and  $p$  that cause some given Regulator Performance Index to take on its minimum value. To solve the Regulator Problem for any given values of  $z$  and  $p$ , any of the three methods discussed previously could conceivably be used, but only Method I lends itself to computer solution.

Although a rather lengthy discussion of the Regulator Problem has been given, there are many questions still unanswered. Some of these questions are raised in Chapter V.

CHAPTER IV  
THE SERVOMECHANISM PROBLEM

The second of the two problems introduced in Chapter II is defined precisely below. This definition of the Servomechanism Problem is again adapted from that of R. J. Leake (1964, p. 14).

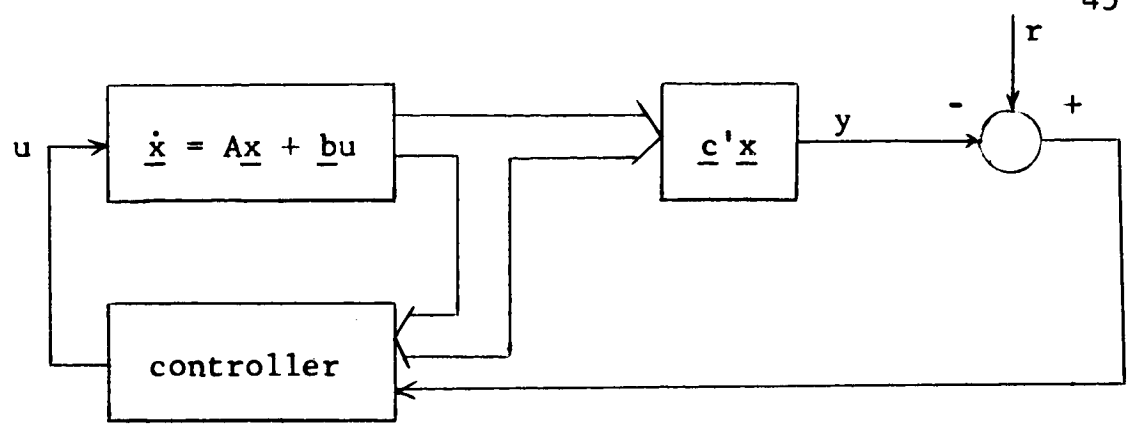
Servomechanism Problem Consider the completely controllable system (2-1) together with the Performance Index

$$J_s = \int_0^{\infty} [(r - \underline{c}'\underline{x})^2 + u^2] dt \quad (4-1)$$

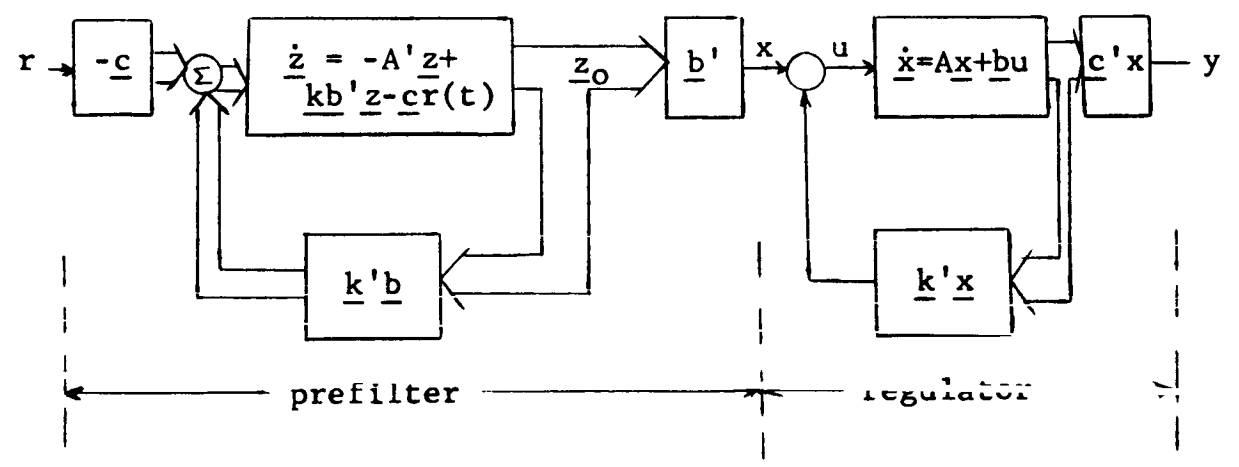
where the function  $r(t)$ , the reference or desired output, is specified to be one such that for some continuous control function  $u$ ,  $J_s$  is bounded.

Assuming arbitrary initial conditions, find a continuous control function  $u$  that minimizes  $J_s$ .

A pictorial interpretation of the Servomechanism Problem is shown in Fig. 7(a). It differs from the Regulator Problem in that the quadratic form  $\underline{x}'Q\underline{x}$  in the integrand of (3-1) is replaced by the square of the difference



a) Problem Interpretation



b) Solution

Fig. 7 The Servomechanism Problem

between the given reference output,  $r(t)$ , and the actual output,  $y(t)$ .

The solution to the Servomechanism Problem, presented in Leake (1964), is repeated here without proof.

Theorem II Let  $r(t)$  be of the form (4-2)

$$r(t) = r_1(t)\exp(-\lambda_1 t) + \dots + r_m(t)\exp(-\lambda_m t)$$

where  $\lambda_1, \lambda_2, \dots, \lambda_m$  are complex numbers with positive real parts and  $r_1(t), \dots, r_m(t)$  are polynomials in  $t$ . Let  $\underline{z}$  denote the particular solution of the differential equation

$$\dot{\underline{z}} = -(A - \underline{bk}')' \underline{z} - \underline{c}r(t) \quad (4-3)$$

where the elements of  $\underline{k}$  are the feedback coefficients of the corresponding Regulator Problem. Then  $\underline{z}$  exists and is unique, and the optimal control function for the Servomechanism Problem for an  $r(t)$  of the form (4-2) is given by

$$u = r_0(t) - \underline{k}' \underline{x} \quad (4-4)$$

where

$$r_0(t) = \underline{b}' \underline{z}(t) \quad (4-5)$$

The structure of the optimal servomechanism, as deduced from Theorem II and Fig. 7(b), consists of two parts: the prefilter and the regulator. The form of the

prefilter depends on the reference signal  $r(t)$ , so that to calculate  $\underline{z}(t)$  it is necessary to know  $r(t)$  in the interval  $(0, \infty)$ . The regulator is found by solving the associated Regulator Problem; this is an important observation because it implies that results obtained for either problem (Servomechanism or Regulator) will apply, in part, to the remaining problem.

Utilizing Theorem II and Kalman's Equation, frequency-domain equations can be derived for the two parts of the optimal servomechanism. For this purpose let the symbol  $\{ \}^+$  denote an extraction of the multiplicative factor containing the left half plane poles and zeroes and let the symbol  $[ ]^+$  denote the sum of those terms in the partial fraction expansion of the enclosed quantity which contain the LHP poles.

Recall that for a completely observable system the regulator portion of the compensated system is stable. Using the model shown in Fig. 4 of Chapter III the poles and zeroes of  $1 + GH(s)$  in the equation

$$\frac{y(s)}{r(s)} = \frac{G(s)}{1 + GH(s)} = M_1(s) \quad (4-6)$$

must lie in the left half plane. Now writing Kalman's Equation as



$$\begin{aligned} [1 + GH(s)] [1 + GH(-s)] &= 1 + G(s)G(-s) \\ &= \{1 + G(s)G(-s)\}^+ \{1 + G(s)G(-s)\}^- \end{aligned} \quad (4-7)$$

it is apparent that

$$1 + GH(s) = \{1 + G(s)G(-s)\}^+$$

Substituting (4-7) into (4-6) gives

$$M_1(s) = \frac{G(s)}{\{1 + G(s)G(-s)\}^+} \quad (\text{regulator}) \quad (4-8)$$

This is the frequency-domain expression for the regulator portion of the optimal Servomechanism. Using (4-3) and the fact that  $\underline{z}$  is the particular solution of that equation, an expression of the prefilter can be derived (Leake 1964). The result is

$$M_2(s) = \frac{1}{r(s)} [M_1(-s)r(s)]^+ \quad (\text{prefilter}) \quad (4-9)$$

Equations (4-8) and (4-9) present a relatively new solution to an old problem. The Servomechanism Problem was first solved by using Parseval's Theorem (Chang 1961); Chang derives an expression for  $M_1(s)M_2(s)$ , the overall transfer function. This older treatment, however, is not able to distinguish between the regulator and the prefilter portions of the system; furthermore, the older results are valid only for zero initial conditions. On the other hand, Theorem II shows clearly that the optimal control system

consists of a prefilter, whose output is the filtered reference signal, and a regulator, which utilizes feedback from all the state variables. The feedback coefficients are independent of the reference signal and the system is optimal for arbitrary initial values of the state variables.

For higher order systems the spectral factorization required in (4-8) is an obstacle in the design procedure, unless graphical techniques are used. It is necessary to find the LHP factors of

$$1 + G(s)G(-s) = 0 \quad (4-10)$$

This is a root locus problem, and as such is familiar to engineers acquainted with conventional control theory. Starting with the poles and zeroes of  $G(s)G(-s)$  the locus of the roots of (4-10) are sketched, and the LHP factors are obtained. It is then a simple matter to compare the expression for  $M_1(s)$  calculated from (4-8) with the expression

$$\frac{G(s)}{1 + GH(s)}$$

in order to evaluate the feedback coefficients.

The following example illustrates the two-part structure of the optimal system and the use of root locus techniques.

Example 6

Consider the uncompensated system of Example 3, Chapter III, where

$$G(s) = \frac{10}{(s)(s+1)}$$

Let the output be  $x_1$ , so that  $H(s)$  is given by

$$\begin{aligned} H(s) &= \frac{k_1 x_1(s) + k_2 x_2(s)}{x_1(s)} \\ &= k_1 + k_2(s+1) \end{aligned}$$

and Performance Index is given by

$$J_s = \int_0^{\infty} [(r - x_1)^2 + u^2] dt$$

The Root Locus Plot of  $1 + G(s)G(-s)$  is shown in Fig. 8.

From the plot,

$$\{1 + G(s)G(-s)\}^+ = (s + 2.3 + j2.2)(s + 2.3 - j2.2)$$

Substituting into (4-8) gives

$$\begin{aligned} \frac{G(s)}{\{1 + G(s)G(-s)\}^+} &= \frac{\frac{10}{(s)(s+1)}}{(s + 2.3 + j2.2)(s + 2.3 - j2.2)} \\ &= \frac{10}{s^2 + 4.6s + 10} \end{aligned}$$

The above expression for  $M_1(s)$  is to be compared with the expression obtained by using (4-6), namely

$$\frac{G}{1 + GH} = \frac{10}{s^2 + (10k_2 + 1)s + 10(k_1 + k_2)}$$

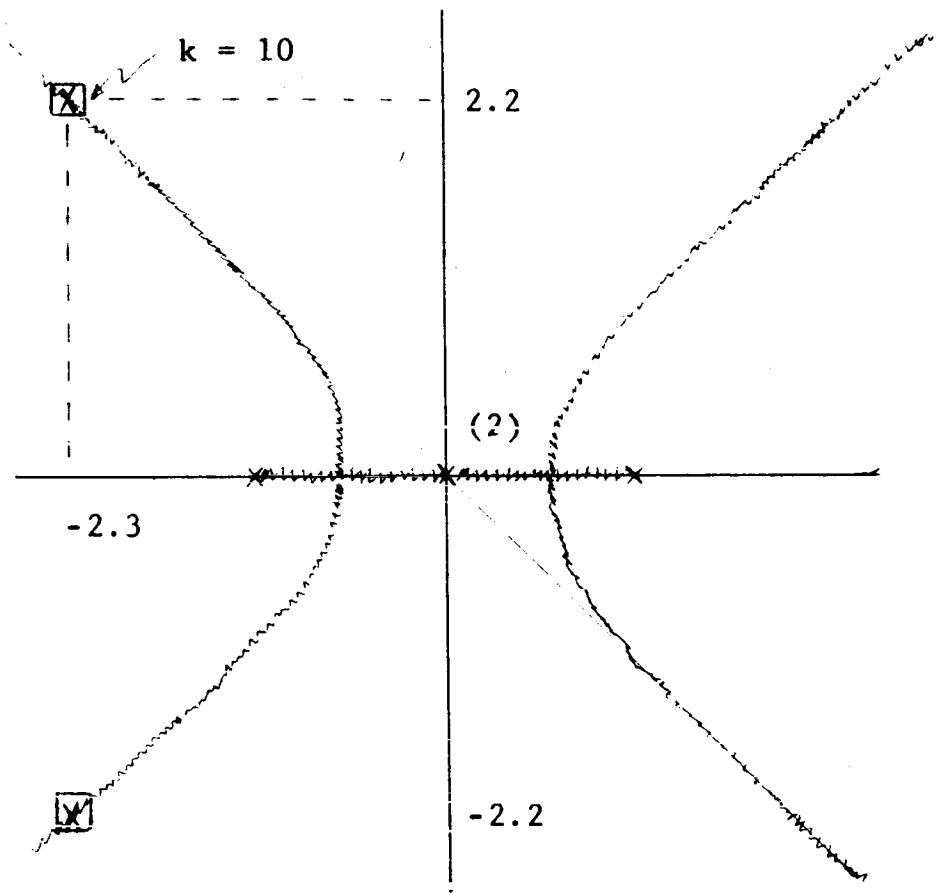


Fig. 8 Root Locus Plot  
for Example 5

Equality of the two expressions for  $M_1(s)$  requires  $k_1 = .64$  and  $k_2 = .36$ . These values of the feedback coefficients are the same as those found in Example 3.

To complete this example the prefilter portion of the optimal system will now be calculated. Let the reference output  $r(t)$  be

$$r(t) = 1 - \exp(-t)$$

so that

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

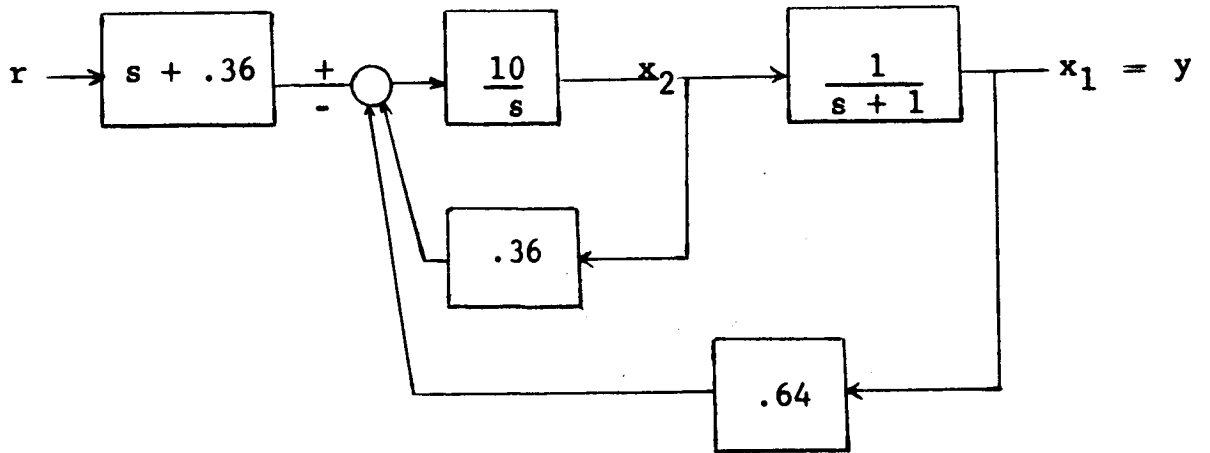
Using (4-9)

$$\begin{aligned} M_2(s) &= (s)(s+1) \left[ \frac{10}{(s^2 - 4.65 + 10)(s)(s+1)} \right] + \\ &= (s)(s+1) \left[ \frac{1}{s} - \frac{.64}{s+1} \right] \\ &= s + .36 \end{aligned}$$

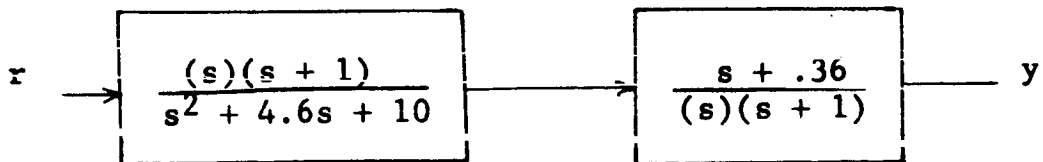
The overall transfer function is given by  $M_1(s)M_2(s)$ :

$$\frac{y(s)}{r_o(s)} = \frac{10(s + .36)}{s^2 + 4.65 + 10}$$

The compensated system is shown in Fig. 9(a). If the older method of solution in the frequency domain were applied to this example, then the same overall transfer function would have been found. But then the designer would have had no aid in determining the best way of



a) Optimal Realization of the Compensated System



b) Alternate Realization

Fig. 9 The Compensated Systems of Example 5

implementing that overall transfer function. If he chose to realize the optimal system as Fig. 9(b), for example, then his design would be optimal only for zero initial conditions, whereas Fig. 9(a) is optimal for any initial conditions on the state variables  $x_1$  and  $x_2$ .

It is well to note that the procedure for calculating the regulator portion of the optimal servomechanism can be applied to the Regulator Problem when the Performance Index  $J_r$  has the form

$$J_r = \int_0^{\infty} (\underline{x}' \underline{c} \underline{c}' \underline{x} + u^2) dt \quad (4-11)$$

Thus the expression for  $M_1(s)$  given in (4-8) can be used along with the root locus techniques to find a completely determined expression for the overall transfer function of the Regulator Problem; this expression can then be compared with the one obtained from the knowledge of the structure of the optimal system and the feedback coefficients can be evaluated. This sequence of steps for calculating the feedback coefficients is very similar to Method III and has already been illustrated in Example 6.

Some insight into the relation between the Regulator Problem and the Servomechanism Problem can be obtained by finding the prefilter for a fixed plant with at least one

pure integration and a step-function reference input (Leake 1964). With an integration in  $G(s)$ , (4-6) yields the result

$$M_1(0) = 1 \quad (4-12)$$

Using (4-12) and (4-9), with  $r(s) = 1/s$ , it follows that

$$\begin{aligned} M_2(s) &= \frac{1}{r(s)} \left[ r(s)M_1(-s) \right]^+ \\ &= s \cdot \frac{1}{s} \\ &= 1 \end{aligned}$$

so that the overall transfer function of the compensated system becomes

$$\begin{aligned} M(s) &= M_1(s)M_2(s) \\ &= M_1(s) \end{aligned} \quad (4-13)$$

From (4-13) it can be concluded that the regulator portion of the optimal Servomechanism is by itself the complete optimal solution when the reference input is a step function and the fixed plant has at least one pure integration. This is an important practical result.

While the treatment of the Servomechanism Problem has been less complete than that of the Regulator Problem the two are so closely related that this manner of presentation is justified.



## CHAPTER V

### CONCLUSIONS

Based on the research reported in Chapter III, it appears that the Regulator Problem has been studied thoroughly. The effect of feeding back all the state variables on the poles and zeroes of the system is known, and several methods are available for calculating the feedback coefficients for a given design problem; the relation between the parameter optimization problem and the Regulator Problem is clear.

However, the discussion of the Regulator Problem is incomplete for several reasons. First, there is no good way of picking the matrix  $Q$  in the expression for the Regulator Performance Index. Second, the limitations of state-variable feedback are not fully known; e.g., can the designer be sure that he will always get a satisfactory design by feeding back all the state variables? Third, the connections with conventional, frequency-domain design techniques have not been established. It should be noted that these three comments have been frequently made in discussions of the relative merits of state variable techniques and conventional techniques.

The rebuttal to the first criticism has usually made use of the concept of modeling. In brief, the model is the form that the compensated system would have to assume in order to achieve the lowest possible value of the chosen Performance Index. Frequently, design constraints permit only enough freedom to achieve a value which is greater than the lowest possible value. Modeling is useful when the choice of the model specifies the Performance Index, since then the designer has a means of selecting the Performance Index. For the Regulator Problem very little attention has been given to development of models, perhaps because the integrand of  $J_r$  in (3-1) becomes quite complicated when the substitution  $u = \underline{k}'\underline{x}$  is made. Some work has been done for the case where the integrand is  $\rho \underline{x}'\underline{Q}\underline{x} + u^2$ ; as  $\rho \rightarrow \infty$  the poles of the optimal system approach a Butterworth configuration (Kalman 1964, p. 58).

The second criticism - that of the lack of knowledge of the limitations of state variable feedback - has been partly clarified by showing that the use of state variable feedback allows complete control over the poles of the system but does not affect zeroes. What remains to be given is a more complete treatment of the techniques

for including zeroes in a meaningful way in the Regulator Problem.

The third criticism is of particular importance for the Regulator Problem because the form of the optimal system is known. This structure could serve as a starting point for applying, for example, the minor loop design techniques to the Regulator Problem. The subjects of sensitivity and steady-state error constants also remain to be considered.

The Servomechanism Problem was given a briefer treatment than the Regulator Problem because once the latter is understood, the former is easily grasped. All of the comments given above apply equally well to the Servomechanism Problem. In one sense the Servomechanism Problem has a closer appeal to automatic control engineers since the reference input is an explicit part of the problem. Note, however, that these engineers usually use the response of the system to a step input as a reliable guide to the merit of the design; for this particular reference input the Servomechanism Problem reduces to the Regulator Problem, as was shown in Chapter IV.

In conclusion, the two basic problems considered in the thesis have been carefully defined and an attempt

has been made to treat each one as an individual problem, to relate each one to the other, to relate both to other design techniques, and to point out the areas where further investigation is required.

## REFERENCES

- Brockett, R. W. Poles, Zeroes and Feedback: State Space Interpretation. IEEE Transactions on Automatic Control, April 1965.
- Chang, S. S. L. Synthesis of Optimum Control Systems McGraw Hill, 1961.
- Gibson, J. E. Nonlinear Automatic Control. McGraw Hill, 1963.
- Kalman, R. E. Mathematical Description of Linear Dynamical Systems. SIAM Journal on Control, 1963.
- Kalman, R. E. When Is a Linear Control System Optimal? Journal of Basic Engineering, March 1964.
- Leake, R. J. Analytical Controller Design and the Servomechanism Problem. Syracuse University, Electrical Engineering Department Technical Report TR-64-4, July 1964.
- Leake, R. J. Return Difference Bode Diagram for Optimal System Design. IEEE Transactions on Automatic Control, July 1965.