

Negative microfilm
received from university

65-5920

KUO, Marshall Chuan Yung, 1928-
THE APPLICATION OF FUNCTIONAL
ANALYSIS TO SOLVE A CLASS OF
LINEAR CONTROL PROBLEMS.

The University of Michigan, Ph.D., 1964
Engineering, electrical

University Microfilms, Inc., Ann Arbor, Michigan

**THE APPLICATION OF FUNCTIONAL ANALYSIS TO SOLVE
A CLASS OF LINEAR CONTROL PROBLEMS**

**by
Marshall Chuan Yung Kuo**

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in The
University of Michigan
1964**

Doctoral Committee:

**Professor Louis F. Kazda, Chairman
Associate Professor Kuei Chuang
Professor Louis J. Cutrona
Professor Albert E. Heins
Associate Professor William A. Porter**

ACKNOWLEDGMENTS

The author wishes to express his thanks to the assistance and encouragement received from the faculty of the Department of Electrical Engineering and the Navigation and Control Systems Laboratory of the University's Institute of Science and Technology. The continuous aid given by Professor L. F. Kazda and Professor W. A. Porter is especially acknowledged.

The suggestions, comments and criticisms of Professor L. F. Kazda, L. J. Cutrona, A. E. Heins, Kuei Chuang, W. A. Porter and G. J. Minty are sincerely appreciated.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	11
LIST OF TABLES	vii
LIST OF SYMBOLS	viii
LIST OF FIGURES	xiv
Chapter	
I. INTRODUCTION	1
1.1 Optimal Control Process	1
1.1.1 The Concepts of Optimum Control and Performance Index.	1
1.1.2 Types of Optimum Control Problems	4
1.1.3 Mathematical Approach Available to Solve Optimal Control Problems	8
1.1.4 Historical Background of the Optimal Control.	8
1.2 Introduction of the Method of Steepest Descent	11
1.3 Introduction of the Concept of Reliability.	14
1.4 Problems and Objectives of this Dissertation	16
II. MINIMUM ENERGY PROBLEMS	18
2.1 Introduction	18
2.2 Problem 1	20
2.2.1 Abstract Theory	20
2.2.2 Examples	25
Example 1	25
Example 2	26
2.2.3 Physical Applications	27
Example 1	27
Example 2	33
2.3 Problem 2	37
2.3.1 Abstract Theory	38
2.3.2 Example	41
2.4 Problem 3	43

	Page
2.4.1 Abstract Theory	43
2.4.2 Examples	49
Example 1	49
Example 2	52
Example 3	57
2.5 Problem 4	71
2.5.1 Abstract Theory	72
2.5.2 Examples	75
Example 1	75
Example 2	78
Example 3	79
III. COMPARISON WITH OTHER OPTIMIZING TECHNIQUES	83
3.1 Abstract Operator Formulation	83
3.2 Dynamic Programming Approach	86
3.3 Pontryagin's Approach	88
3.4 Calculus of Variations	90
3.5 A Method for Evaluating the Optimal Performance Index Given by	
$J^0 = \frac{1}{2} \int_{t_0}^{t_f} \{ [x, Qx] + [u^0, Ru^0] \} ds$	91
3.6 The Equivalent Expression of Optimal Control Function and Optimal Performance Index	95
3.6.1 Example 1	95
3.6.2 Example 2	97
IV. CONCLUSIONS AND SUGGESTIONS FOR FUTURE EXTENSION	102
4.1 Conclusions	102
4.2 Suggestions for Future Work	105
APPENDIX AI. REVIEW OF BASIC TERMINOLOGY IN FUNCTIONAL ANALYSIS.	108
APPENDIX AII. THE DYADIC REPRESENTATION	115
APPENDIX AIII. THE OPERATOR REPRESENTATION OF LINEAR DYNAMIC SYSTEM	118
AIII-1 Continuous Linear System	118
AIII-2 Discrete Linear System	123

	Page
APPENDIX AIV. PROOF THAT $(Tv)(t)$ DEFINED BY	
$(Tv)(t) = \Phi(t, t_0)x_0 + \int_{t_0}^t \Phi(t, s)B(s)u(s)ds$	
IS A LINEAR AND BOUNDED OPERATOR . . .	128
APPENDIX AV. PROOF THAT F^t DEFINED BY	
$F^t u = \int_{t_0}^t \Phi(t, s)B(s)u(s)ds$	
IS BOUNDED FOR FIXED t IN THE INTERVAL τ , WHERE	
$F^t : L_2^m(\tau) \rightarrow R^n$	130
APPENDIX AVI. A PROOF OF	
$F^*x = B^*(t)\Phi^*(t_0, t) \int_t^{t_f} \Phi^*(s, t_0)x(s)ds,$	
$t \in \tau$	133
APPENDIX AVII. A PROOF OF	
$(F^{t_f})^*x(t_f) = B^*(t)\Phi^*(t_0, t)$	
$\Phi^*(t_f, t_0)x(t_f)$	135
APPENDIX AVIII. A PROOF OF $\mathcal{F}^{-1} = [\langle f_1^*, f_j^* \rangle]$	136
APPENDIX AIX. A PROOF OF	
$\ \delta x\ ^2 \leq M^2 \int_{t_0}^{t_f} [\bar{x}^* \delta A^* \delta A \bar{x}]dt$	138
APPENDIX AX. THE METHOD OF STEEPEST DESCENT	140
AX-1 The Solution to the Operator Equation $y = Kx$	142
AX-2 Minimization of $z = f(x)$ With the Inequality Constraint $g(x) \leq 0$	147
APPENDIX AXI. A PROOF OF $H(F)^{\perp} = VH(F^*)$	161

	Page
APPENDIX AXII. A PROOF OF LEMMA (1) IN SECTION	
2.4.1	162
LIST OF REFERENCES	164

LIST OF TABLES

Table	Page
1. Some Numerical Comparisons of J^* Obtained from Equations (2.4.2-60) and (2.4.2-62). . . .	70
2. Comparison of Optimization Techniques	101

LIST OF SYMBOLS

$A(t)$	=	$n \times n$ time varying matrix
$B(t)$	=	$n \times m$ time varying matrix
C	=	The scalar cost function
$C(a,b)$	=	space of continuous functions defined on the interval $[a,b]$
$E(.)$	=	The expectation of the random variable $(\cdot)$
F	=	A bounded linear operator mapping of H_1 into H_2
F^t	=	A bounded linear transformation which maps H onto R^n and is defined by $F^t u = \sum_{i=1}^n \phi_i(t) \langle f_i, u \rangle$
F^S	=	A parametrized family of linear transformations each of which is a mapping of H onto R^n
F^*	=	The adjoint operator of F
$F^\dagger$	=	The pseudoinverse of F
G^t	=	A bounded linear operator which maps $H(F)$ onto R^n and is defined by $G^t(u, Fu) = F^t u = x(t)$
H	=	Hilbert function space
$H(F)$	=	The graph of operator F
H_p	=	The projection of $H_1 \times H_2$ on $H(F)$
I	=	The identity matrix

J	= The index of performance
J^o	= The optimal index of performance
K	= A bounded linear operator
$L_2(t_o, t_f)$	= L_2 space in the closed interval $[t_o, t_f]$
M	= A linear manifold spanned by the vector $\left\{ r_1 \right\}_1^n$
$M^\perp$	= The orthogonal complement space of M
N_{F^t}	= The null space of F^t
O	= The complement of the null space of G^t which is defined by $O = \left\{ (u, Fu) : (R(t) + F^*Q(t)F) u \in M \right\}$
$P(t)$	= A symmetric matrix which is satisfied by the matrix Riccati equation
P_L	= $(1 - \nabla g \frac{\nabla g^*}{\ \nabla g\ ^2})$ a projection operator de- fined on R^n with range in the subspace L , where ∇g^* is the transpose of ∇g
P_s	= The chord formed by the shadow of ∇f on the surface $g(x)=0$
Q	= The complement of the null space of G^t which is defined by $Q = \left\{ (u, Fu) : (I + F^*F)u \in M \right\}$
$Q(t), R(t)$	= Positive definite matrices
R^n	= Euclidean space
S	= The terminal set defined by $S = X \times T$
S_u	= The space formed by the projection of $H(F)$ on H

- T = A bounded linear operator mapping of $R^n \times U$ into X
- T^t = A bounded linear transformation defined by $T^t v = (Tv)(t)$
- T_e = A scalar terminal error measure
- T_1^t = A mapping of R^n into R^n
- T_2^t = A mapping of R^n into R^n
- T_{12}^t = $T_1^t \oplus T_2^t$ a mapping of $R^n \times U$ into R^n
- U = The input space of control system
- U_x = The set defined by $U_x = \{u : u \in U, F^t u = x(t)\}$
- U_p = $L_{p_1} \times \dots \times L_{p_m}$ the cartesian product of the Banach space $L_{p_i} (i = 1 \dots m)$
- U^*, X^* = The dual space of U and X respectively
- X = The output space of control system
- a = Random variable
- $\bar{a}$ = The mean value of the random variable "a"
- $\langle e_1^*$ = The linear functional defined on R^n
- $f(a)$ = The probability density function of a
- $f(x^{(j)})$ = The value of the function f evaluated at $x^{(j)}$
- $\langle g_1$ = $\langle (e_1^*, f_1)$ a linear functional defined on $R^n \times U$
- $\langle h_1$ = $\langle (\phi_1^+, f_1)$ a linear functional defined on $R^n \times U$

- k = A scalar factor for the determination of the size of step in the method of steepest descent
- $l_2(\delta)$ = l_2 space defined on $\delta = [t_0, t_1, \dots, t_f]$
- $[t]$ = $\text{Max.}_{t_k \in \delta} \{k : t_k \leq t\}$, $[t] = k$ if and only if $t \in [t_k, t_{k+1}]$
- u = The control function
- u° = The optimal control function
- x = The state of the dynamic system
- x_0, x_f = The initial and final value of x at $t = t_0$ and $t = t_f$
- δx = The deviation of x from its nominal trajectory $\bar{x}$
- $\|x\|$ = The norm of x
- $\in$ = Belong to or an element of
- τ = Closed time interval $[t_0, t_f]$
- ν = Time parameter set where
 $\nu = \tau = [t_0, t_f]$ in continuous case
 $\nu = \delta = [t_0, t_1, \dots, t_f]$ in discrete case
- ϕ_i = The i -th column of the matrix $\Phi(t_0, t_f)$
- ϕ_i^+ = The dual vector of ϕ_i
- β = A constant that is determined by the terminal value of the state of system
- $\mathcal{F}$ = A matrix defined by $[<f_i, f_j>]$ which in matrix form is expressed by

$$\mathcal{F} = \int_{t_0}^{t_f} \Phi(t_0, S) B(S) B^*(S) \Phi^*(t_0, S) dS$$

In continuous case

$$\mathcal{F} = \sum_{\alpha=0}^{[t_f]-1} h_{\alpha} \Phi(t_0, t_{\alpha+1}) B(t_{\alpha}) B^*(t_{\alpha}) \Phi^*(t_0, t_{\alpha+1})$$

In discrete case

α

= A constant that is determined by $\mathcal{F}^{-1} \beta$

where $\mathcal{F}^{-1}$ is inverse of $\mathcal{F}$

$\mathcal{R}$

= $\{x : g(x) \leq 0\}$

$\varphi_u(t, x_0, t_0)$

= The solution of the system given by $\dot{x}(t) =$

$A(t)x(t) + B(t)u(t)$, $x(t_0) = x_0$.

Γ

= $\{ \xi : \xi_1^2 + \dots + \xi_n^2 \leq \epsilon^2 \}$

$\mathcal{H}(x)$

= $\left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right]$ a Hessian matrix

$\mathcal{H}$

= A matrix defined by $[<h_i, h_j>]$

$\lambda(t)$

= The time-dependent Lagrange's multiplier

$\psi(t)$

= $(\psi_1(t), \psi_2(t) \dots \psi_n(t))$ nonzero continuous vector function

$\Phi(t, t_0)$

= The transition matrix of the system given by

$\dot{x}(t) = A(t)x(t) + B(t)u(t)$, $x(t_0) = x_0$.

$\Psi(t, t_0)$

= The transition matrix of the adjoint system given by $\dot{\lambda}(t) = -A^*(t)\lambda(t)$

$\textcircled{H}(t, t_0)$

= The grand transition matrix of the system given by

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\lambda}(t) \end{bmatrix} = \begin{bmatrix} A(t) & B(t)R^{-1}(t)B^*(t) \\ Q(t) & -A^*(t) \end{bmatrix} \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix}$$

- τ = The time set
- $\nabla_{\mathbf{x}} J$ = col. $\left[\frac{\partial J}{\partial x_1}, \dots, \frac{\partial J}{\partial x_n} \right]$ (the gradient of J)
- $\Delta \mathbf{x}(t_k)$ = $h_k^{-1} [\mathbf{x}(t_{k+1}) - \mathbf{x}(t_k)]$ where $h_k = t_{k+1} - t_k$
- $[u, v]$ = The inner product of elements u and v
- $\langle f, u \rangle$ = The functional f acting on the vector u

LIST OF FIGURES

Figure		Page
2.2.3-1	Second-Order Position Control Servomechanism	29
2.2.3-2	First-Order System with Sampler and Zero Order Hold.	33
2.4.2-1	A Feedback Control System with Random Variable "a" and Unknown Parameter $\xi = (\xi_1, \xi_2)$	63
4.2-1	Adaptive Optimum System.	106
AVI-1	Domain of Integration.	133
AX-1	A Geometric Interpretation of the Method of Steepest Descent in the Minimization of $z = f(x_1, x_2)$	141
AX-2.1	A Geometric Meaning of P_L and P_S	154
AX-2.2	A Geometric Interpretation of the Method of Steepest Descent in the Minimization of $z = f(x_1, x_2)$ with Constraint $g(x_1, x_2) = 0$	155

CHAPTER I

INTRODUCTION

1.1 Optimal Control Process

1.1.1 The Concepts of Optimum Control and Performance Index

The fundamental mathematical problem in the design of a control system is the specification of the control law.¹ They can be realized in various ways depending on the will of man. As the classical design techniques such as the Nyquist plot, root locus method, and so forth, have been reached their fruition, the tendency of research work in the area of control system has been toward optimization. This is due to the increasing demand for systems of high performance and to the availability of digital computers which often make feasible the calculations attendant on the design of an optimum control system.

A physical dynamical system to be controlled can often be described by a system of n differential equations or in vector form

$$\dot{x}(t) = f(x(t), u(t), t) \quad (1.1.1-1)$$

¹We denote this functional relationship by $u(t) = k(x(t), t)$ and refer to the function k as the control law. A control law is admissible if $k(x(t), t)$ is admissible.

where

$$x = (x_1 \dots x_n) \in X$$

$$u = (u_1 \dots u_m) \in U$$

t (time) is a scalar in $\mathcal{T}$

$x(t)$ in equation 1.1.1-1 is called the state of the dynamic system, it is a minimum set of numbers which, specified at any given time, suffice to determine completely the future evolution of the system, provided the future forces acting on the system are known. The state of the system is a (finite dimensional) vector, $x_i(t)$ ($i=1,2, \dots n$), the components of $x(t)$, are called the state variables which characterize the control process. For example, a system of N particles free to move in 3-dimensional space has a state vector containing $6 N$ components, composed of $3 N$ position and $3 N$ velocity coordinates. In some cases, the dimension of a state vector may be infinite, as in partial differential equations. In equation 1.1.1-1 $u(t)$ is called the control function which determines the course of the process. For example in a mechanical control problem the control function $u(t)$ determines the position of a machine's controllers. In most optimal control problems, $u(t)$ is assumed to be an admissible function. Definition: $u(t) \in U$ is called an admissible function if U satisfies the following three conditions (68).

(1) $u(t)$ is measurable and bounded for $t_0 \leq t \leq t_f$

(2) if $u(t)$, $t_0 \leq t \leq t_f$ is an admissible control, v is an arbitrary point of U and t' and t'' are numbers such that

$$t_0 \leq t' \leq t'' \leq t_f$$

The control $u_1(t)$, $t_0 \leq t \leq t_f$ defined by the formula

$$u_1(t) = \begin{cases} v & \text{for } t' \leq t \leq t'' \\ u(t) & \text{for } t_0 \leq t \leq t' \text{ or } t'' \leq t \leq t_f \end{cases}$$

is also admissible.

(3) If the interval $t_0 \leq t \leq t_f$ is broken up by means of a subdivision points into a finite number of subintervals, on each of which the control $u(t)$ is admissible, then this control is also admissible on the entire interval $t_0 \leq t \leq t_f$. A control obtained from an admissible control $u(t)$, $t_0 \leq t \leq t_f$, by a translation in time is also admissible.

Example: The set of all piecewise continuous controls is a class of admissible controls.

In this dissertation the dynamic system is assumed to be described by a set of linear differential equations

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \text{ with } x(t_0) = x_0 \quad (1.1.1-2)$$

where

$x(t)$ is an n dimensional vector in X

$u(t)$ is an m dimensional vector in U

$A(t)$ is $n \times n$ time varying matrix

and $B(t)$ is $n \times m$ time varying matrix

By the equations of motion of a dynamic system, is meant a rule which specifies how the state of the system at a given time is transformed into another state in the future. This process is referred as state transition. The solution of homogeneous equation to 1.1.1-2 denoted by $\Phi(t, t_0)$, there-

fore, is called the transition matrix which characterizes the motion of $x(t)$ from t_0 to t .

The problem of optimal control is to find the present control law $u(t)=K(t,x(t))$ and to force a dynamic process governed by equation 1.1.1-2 to perform in some best manner. To make the problem meaningful it is necessary to define a quantitative index of performance. Mathematically it may be assumed to have the following form (37, 38).

$$J = T_e(x_f, t_f) + \int_{t_0}^{t_f} C(\varphi_u(s, x_0, t_0), u(s), s) ds \quad (1.1.1-3)$$

where C is a scalar cost function

T_e is a scalar terminal error measure

$\varphi_u(s, x_0, t_0)$ is the solution to 1.1.1-2, x_0 and x_f are respectively the initial and final value of $x(t)$ at $t = t_0$ and $t = t_f$.

The general optimum control problem that we treated here can be stated as follows:

Given the initial state x_0 of the system given by 1.1.1-1 to find a control function $u(t)$ such that $(x_f, t_f) \in S$, where S is a terminal set in $X \times T$, and J takes on the least possible value.

1.1.2 Types of Optimum Control Problems

(A) Terminal Control: Given initial state x_0 , choose an admissible $u(t)$ to drive x_f as close as possible to a specified state γ at specified t_f . For this

$$T_e = \|r - x_f\|$$

$$C = 0$$

$$S = X \times t_f$$

A class of problems of linear terminal control can be found in references (9), (16) and (72).

- (B) Minimal Time Control: Given x_0 , choose an admissible $u(t)$ such that $x_f = r$ in shortest possible time.

For this

$$T_e = 0$$

$$C = 1$$

$$S = r \times \tau$$

A number of problems of minimum time control have appeared in the literature (2), (7), (22), (24), (51), (64) and (65). The time optimal control can be obtained as an explicit function of the state for linear second order systems (34) and for some linear third order system by means of switching surfaces (80).

- (C) Pursuit Problem: Given x_0 and the moving target state $r(t)$, choose $u(t)$ to drive the state x_f to $r(t)$ as soon as possible. This is a generalization of the minimal-time control. This gives

$$T_e = 0$$

$$C = 1$$

$$S = \{ (r(t), t) : t \in \tau \}$$

For example D. L. Kelendzheridge, "On The Theory of Optimal Pursuit," dealt with this type of problem (43).

- (D) Regulator Problem: Given x_0 , choose $u(t)$ to minimize some cost function C , which depends of u , at specified time t_f .

$$\begin{aligned}T_e &= 0 \\C &= C(u(t)) \\S &= X \times t_f\end{aligned}$$

An example of a regulator problem can be found in the reference (46).

- (E) Servomechanism Problem: Given x_0 and desired $\gamma(t)$, $t \in T$, choose $u(t)$ to minimize a measure of error $\gamma(t) - g(x, t) = e(t)$ at specified time t_f , where $g(x, t)$ is a known function of x . This gives

$$\begin{aligned}T_e &= 0 \\C &= C(e(t)) \\S &= X \times t_f\end{aligned}$$

Several papers of formal solutions of servomechanism problems have appeared in the literature (30, 62).

- (F) Minimum Effort Control: Given x_0 , choose $u(t)$ to drive x_f to a specified state γ at specified time t_f with minimum control effort. For this

$$\begin{aligned}T_e &= 0 \\C &= C(\|u\|^p) \\S &= \gamma \times t_f.\end{aligned}$$

For example, Lucien W. Neustadt (66) considered the case of minimizing the control effort

$$C = \left[\int_{t_0}^{t_f} \sum_{j=1}^m |u^j(t)|^p dt \right]^{\frac{1}{p}}$$

and determined the minimum effort and optimal control function $u(t)$ by using the method of steepest descent.

- (G) Minimum Fuel Problem: Given x_0 , choose $u(t)$ to drive x_f to state γ at specified t_f with minimum fuel. This gives

$$T_e = 0$$

$$C = (\|u\|)$$

$$S = \gamma \times t_f$$

Previous work concerning the minimum fuel problem can be found from the papers (4), (5) and (7).

- (H) Minimum Energy Problem: Given x_0 choose $u(t)$ to drive x_f to state γ at specified t_f with minimum energy

$$T_e = 0$$

$$C = C(x, u) \quad \text{"power" in system}$$

$$S = \gamma \times t_f$$

From (D) to (H), one could also formulate a problem with $T_e \neq 0$, then terminal error will be considered. Of the most recent papers relating to minimum energy control the references (4), (7), (21), (35), (38), (60), (67), and (69) are pertinent to the present

thesis. The theoretical development in sections 2.2.1 and 2.3.1 of Chapter II follows closely the treatment of Porter (69).

1.1.3 Mathematical Approach Available to Solve Optimal Control Problems

The following list several common approaches to the solution of optimum control problems:

- (A) Calculus of variations: (24)
- (B) Dynamic programming: (10, 62)
- (C) Maximum principle: (14), (15), (21), (25), (32) and (75)
- (D) Measure theory: (51)
- (E) Function space approach: (4), (7), (9), (35) and (69)

1.1.4 Historical Background of the Optimal Control

The optimal control was begun in Russian by Ivachnenko (1948) and Feldbaum (1949), in American by McDonald (1950), and in England by Uttley and Hammond. These early papers dealt with the relay control system with simple linear processes of second order and with the problem of minimizing the settling time after the application of a step input. Hopkin made further contribution along these lines. In 1951 the earliest work on an extremal sensing system of the optimal control type was developed by Drapper and Li. Bushaw, Lasalle, and Rose (1953) generalized the theory for linear processes. These reports were unfortunately not published. Some of Bushaw's work was reported in the book by Tsien (1954). Bogner and Kazda (1954)

again treated higher order systems and this approach was further developed for third order system by Hopkin and Iwana (1956). In the same year (1956) Kuba and Kazda extended further to the system of fourth order by selecting a continuous response curve to approximate the parabolic phase plane trajectories of the ideal switched system. Bellman (1956) developed the theory of dynamic programming which is a general technique of optimization using a digital computer. One drawback of this technique is the memory required for a multidimensional vector x of higher order. For higher dimensions, the required memory quickly rises to astronomical proportions, and a direct application of dynamic programming becomes impractical. R. V. Gamkrelidze, V. G. Boltyanskii, and L. S. Pontryagin (1956) developed a very interesting theory known as maximum principle. This result was particularly interesting because it stated properties of the optimal solution for every general nonlinear dynamic system. In 1959 Desoer solved the minimum settling time by use of calculus of variations and in 1961 he also derived the maximum principle by use of the principle of optimality. The derivation was particularly simple and did not require $\Omega(t)$ (the set in which $u(t)$ lies) to be convex. Giancarlo L. Collina and Peter Dorato (1962) applied the Pontryagin's maximum principle to the solution of linear control systems with quadratic form of index of performance and no constraints on the plant control signals. About the time from 1960 to 1962 a series of papers relating to the

optimal control problem were published by Kalman. Some of his papers discussed the stochastic process. The stochastic optimal control problem was originated by Bellman (1957) and further developed by Bellman (1958), Fuller (1960) and Wonham (1962). In the period starting in about 1959 several papers of minimum energy problem have been successively published. Recently a technique for solving this minimum energy problem is focused on the method of operator formulation and functional analysis. For example, in January 1963, M. Athanassiades developed the optimal control for linear time invariant plant with time, fuel and energy constraint. Several months later in July 1963 he extended the problem of minimum time, minimum fuel and minimum energy control for a class of nonlinear norm-invariant systems. James S. Meditch (October 1963) solved a class of linear feedback minimum energy control systems by introducing one additional coordinate of the state vector, namely, x_{n+1} as the minimum energy criterion, i.e., $x_{n+1} = 1/2 \int_0^{t_f} \langle u(s), u(s) \rangle ds$. About the same time W. A. Porter treated the same minimum energy problem by abstract operator formulation from functional space approach with no constraint on $u(t)$.

The body of this dissertation treats the generalized abstract minimum energy problem with an application to the design of both continuous and discrete linear control systems for fixed and variable boundary conditions. Many physical applications are illustrated such as the application of minimum energy theory to the study of open loop system

reliability. Some of the parameters associated with this optimal solution will be determined by the method of steepest descent. For the benefit of readers not familiar with the method of the steepest descent and the concept of the reliability a brief discussion is given here.

1.2 Introduction of the Method of Steepest Descent

The method of steepest descent in Hilbert space was originated by Kantorovich (1948). A number of papers with application of this approximation method to the solution of optimal control problems can be found in the literature (9, 35). As a matter of fact, many practical problems such as the solution of algebraic, differential, integral and other types of minimization problems can be solved by this approximation technique.

We shall give briefly an outline of the steepest descent method in Hilbert space for the minimization of $f(x)$. Consider for example the following: Let X be a Hilbert space and $y=f(x)$ a functional defined on X . Here f may be either linear or nonlinear. Let z and x be arbitrary elements of the space X . The definition of the differential of an abstract function at the point x is then the following:

Definition (strong or Fréchet differential). If a linear functional $h \in X^*$ (conjugate space of X) exists such that with $\alpha(x,z)$ defined by $f(x + z) - f(x) = h(z) + \alpha(x,z)$ (1.2-1)

we have

$$\frac{\|\alpha(x,z)\|}{\|z\|} \rightarrow 0 \text{ as } \|z\| \rightarrow 0, \quad (1.2-2)$$

then $h(z)$ is called the differential of the function $f(x)$ at the point x along the increment z .

In other words, if the expression $df(x,z)=f(x+z)-f(x)$ has an expression as a linear functional "h" on z plus terms of "order higher than one in $\|z\|$," then f is said to have a strong differential at the point x . In a Hilbert space where functionals have the form

$$h(z) = \langle u, z \rangle$$

it is convenient to think of u as the gradient of $f(x)$ at x , denoted by $\nabla f(x)$. Thus the equation

$$df(x,z) = \langle \nabla f(x), z \rangle + \alpha \langle x, z \rangle \quad (1.2-3)$$

is frequently used. Note also that the functional $h = \langle \nabla f$ is evaluated at the point x and is by no means independent of x .

Our problem is assuming that the function $f(x)$ has a unique minimum at $x = x^*$ an unknown point, how does one go about finding x^* ?

The steepest descent technique is applicable when f has a gradient ∇f at every point of X . If this is so then an initial x_0 is chosen. The number $f(x_0)$ and the functional $\langle \nabla f(x_0)$ can be computed. Next the vector $z_0 \in X$ such that $\|z_0\|=1$ and z_0 maximizes $-df(x, z)$ is chosen. The one dimensional manifold $\{kz_0: k \text{ is real}\}$ is called the "direction of steepest descent."

When $X = H$ a Hilbert space from the Schwarz inequality it is clear that the negative maximum of df in 1.2-3 for

$\|z_0\| = 1$ is attained for the unit vector

$$z_0 = - \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|} \quad (1.2-4)$$

The next approximation x_1 will be given by judiciously choosing $k_0 > 0$ and setting

$$x_1 = x_0 + \delta x_0 = x_0 + k_0 z_0 = x_0 - k_0 \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|}$$

that is the initial estimate is adjusted by a step of size k_0 (i.e., a vector of norm k_0) in the direction of steepest descent. Using x_1 the value of $f(x_1)$ and the function $\nabla f(x_1)$ are then recomputed and the procedure is repeated, generating the sequence

$$x_{n+1} = x_n - k_n \frac{\nabla f(x_n)}{\|\nabla f(x_n)\|} \quad \text{where } n = 1, 2, \dots \quad (1.2-5)$$

The main theoretical problems are:

- (1) To find a good way to choose $k_1, \dots, k_n, \dots$
- (2) To show that the sequence $\{x_j\}$ converges to a point making f a local minimum and
- (3) to show that $f(x)$ has global minima on H . In many cases problem (2) can be answered by an application of the contractive mapping theorem and problem (3) frequently involves introducing convexity assumptions on the function f .

It should be noted that the technique of finding z depends on the type of problem under consideration. The step size k_n is either kept small such that the second order term $\alpha(x, \delta x)$ is negligible or in a special case may be adjusted to

speed convergence of the sequence or give the sequence other desirable properties.

1.3 Introduction of the Concept of Reliability

The problem of reliability has become important in the recent years. In the development of large weapons systems it is of prime importance. Without a doubt unreliability has consequences in cost, time wasted, and in certain instances personal and national security. For example, the failure of a home radio would generally cost the amount needed to replace it, however the failure of the similar transistor in a ballistic missile might fail to start, this leads to the subsequent loss of the missile at tremendous cost.

The abstract concept of reliability may mean different things to different people, and therefore it is difficult to have a satisfactory quantitative definition of reliability. Stated simply, reliability is the capability of an equipment not to break down in operation. When an equipment, such as a switch, a diode, or a very complex machine, works well and works whenever called upon to do the job for which it was designed, such equipment is said to be reliable. The satisfactory performance without breakdown while in use and readiness to perform at the desired time are the criteria for equipment reliability.

A measure of an equipment's reliability is the frequency at which failures occur in time. If there are no failures, the equipment is one hundred per cent reliable, if

the rate of failure is low, the equipment is usually still acceptable. If the failure rate is high, the equipment is unreliable. Reliability predicts mathematically the equipment's behavior under expected operating conditions. Reliability expresses in numbers the chance that a piece of an equipment will operate without failure for a given length of time in an environment for which it was designed.

According to RETMA¹ "the reliability is defined as the probability of a device operating its purpose adequately for the period of time intended under the operating conditions encountered."

Measurement or calculation of the adequate performance of a device requires measuring all important performance parameters. As long as these parameters remains within specified limits the equipment is considered to be operating satisfactorily.

In this dissertation we say that a reliable system is one in which system components do not fail during the period of operation, or if they do, the overall system performance is insensitive or invariant under such failures. A particular type failure is studied in this paper. We shall consider those systems in which the random failure is defined as follows:

Definition of failure: Let $\bar{a}$ denote an expected or normal value for a system parameter, a , and $\delta a = a - \bar{a}$ denote the

¹RETMA, Radio Electronics and Television Manufacturers' Association is now known as ETA (Electronics Industries Association).

deviation of "a" from its nominal value. Then if δa is undetected, it is considered to be a failure. We shall consider "a" to be a random quantity that is a random failure. More generally speaking we shall consider a linear time-varying system whose coefficients are function of random variables of known statistical properties, and the system will be assumed to contain an unknown parameter $\xi = (\xi_1 \dots \xi_n)$ which is to be determined.

The system equation is

$$\dot{x}(t) = A(t, a, \xi) x(t) + B(t) u(t), \quad x(t_0) = x_0$$

and the index of performance has the form

$$J = 1/2 E \|\delta x\|^2 + 1/2 \int_{t_0}^{t_f} [u, Ru] ds$$

Where R is a positive definite matrix, the symbol $[u, v]$ denotes the inner product of u and v , and $E \|\delta x\|^2$ is the r.m.s. value of the terminal miss vector due to the system's randomness. Evidently δx is a function of δa . From the above definition of failure and reliability, the lower the occurrence of failure the higher will be the system reliability. Therefore the high reliability of the system under consideration will be attained when the index of performance J takes on the least possible value, as the result of properly choosing the parameter $\xi = (\xi_1 \dots \xi_n)$ and control function $u(t)$.

1.4 Problems and Objectives of this Dissertation

In Chapter II four classes of the abstract minimum energy problem are discussed and many applications to the

optimal control problems are given for the purpose of illustration.

The generalization of Kalman's method for the evaluation of the optimal performance index and comparison of abstract operator formulation with other techniques are studied in Chapter III.

Conclusions will be stated and suggestions for future extension will be made in Chapter IV.

The operator representation of a linear dynamic system, which is utilized in the formulation of abstract minimum energy problems and the determination of related parameters by the method of steepest descent are stated in the Appendix AIII and AX respectively.

The objectives of this dissertation are:

- (1) To generalize the minimum energy problem as initiated by Porter.
- (2) To develop a technique for the computation of optimal quadratic performance in both fixed and variable boundary problems.
- (3) To derive a method of steepest descent with inequality constraint for the minimization problem.
- (4) To apply the abstract minimum energy theory to the study of open loop system reliability, and
- (5) to relate the abstract theory to physical applications.

CHAPTER II

MINIMUM ENERGY PROBLEMS

2.1 Introduction

In this chapter an abstract operator formulation of the minimum energy problems with its applications to the physical systems is studied. It mainly deals with the following four problems:

Problem 1: Let x be an arbitrary element of R^n . Let F^t be a bounded linear transformation of a Hilbert function space H into R^n and is defined by

$$F^t = \sum_{i=1}^n \phi_i(t) \langle f_i$$

with fixed t . Let the set $U_x \subset H$ be defined by

$$U_x = \left\{ u : u \in H, F^t u = x \right\}.$$

From the set U_x find the element u with minimum norm.

The set of range vectors $\{\phi_i\}_1^n$, and the set of functionals $\{\langle f_i \rangle\}_1^n$, are assumed to be linearly independent. Problem (1) can be applied to the solution of minimum energy problem with fixed boundary condition.

It is proved in this chapter that v , the element which minimizes $J(u) = \|u_0 + v\|^2$, is given by $F^t(x + x_0) - u_0$.

Problem 2: Consider the parametrized family $\{F^s : s \in V\}$ of linear transforms each of which is a mapping of the Hilbert function space H onto R^n . Thus for each $s \in V$ the operator F^s

of this family has the form

$$F^s = \sum_{i=1}^n \phi_i(s) \langle f_i^s \rangle$$

where the vectors $\{\phi_i(s)\}_1^n$ and the functionals $\{\langle f_i^s \rangle\}_1^n$ are linearly independent sets in R^n and U^* respectively. Find the vector $x \in R^n$, the parameter $s \in V$ and the input $u \in U$ such that the constraints

$$x = F^s u$$

$$\psi_i(x, s) = 0 \quad i = 1, 2, \dots, m < n, \quad x \in R^n$$

are satisfied and the functional

$$J = g(x, s) + \|u\|^2$$

is minimized. Problem (2) can be applied to the solution of minimum energy problems with movable bound conditions (see Section 2.3.2).

Problem 3: Let H_1 and H_2 be two Hilbert function spaces and F be a transfer operator carrying H_1 into H_2 . Let F^t be the t -th sample of F a mapping of H_1 onto R^n . If U_x is the set defined by

$$U_x = \left\{ u : u \in H_1, F^t u = x \right\}$$

find the element $u \in U_x$ for which $J(u) = \|Fu\|_{H_2}^2 + \|u\|_{H_1}^2$

is minimized.

Problem 4: Let H_1 and H_2 be two Hilbert function spaces and F be a transfer operator carrying H_1 into H_2 . Let F^t be the t -th sample of F . If $\hat{u}$ and $\hat{y}$ denote arbitrary elements of H_1 and H_2 respectively from the set $U_x = \{u : u \in H_1, F^t u = x\}$

find the element which minimizes the functional

$$J(u) = \|Fu - \hat{y}\|_{H_2}^2 + \|u - \hat{u}\|_{H_1}^2.$$

Remarks (1): We define the phrase "Hilbert function space" over a finite interval τ of reals, or $\delta = [t_0, t_1 \dots t_f]$ to be $L_2(\tau)$, or $l_2(\delta)$ respectively.

Remarks (2): Since the range of F is $C(\tau) \cap H_2$, the notation of the "value at a point" of a function in the range makes sense. We define $F^t u = (Fu)(t)$ where Fu is the unique continuous image of u under F . (See Appendix AIII.) In the discrete case, of course, these remarks are superfluous.

2.2 Problem 1

Let x denote an arbitrary element of R^n and let the set U_x be defined by

$$U_x = \left\{ u : u \in U, F^t u = x \in R^n \right\}$$

From the set U_x find the element u with minimum norm.

2.2.1 Abstract Theory

Consider the bounded linear transformation F^t of an arbitrary Hilbert function space U into R^n defined by

$$F^t = \sum_{i=1}^n \phi_i(t) \langle f_i \rangle \quad (2.2.1-1)$$

Let the range vectors $\left\{ \phi_i \right\}_1^n$ and the functionals on U $\left\{ \langle f_i \rangle \right\}_1^n$

comprise linear independent sets, then F^t is a linear transformation onto R^n . The term "minimum energy" is derived from the fact that many physical problems involving energy can be

formulated such that the signals $u(t)$ involved are considered as elements in a Hilbert function space of functions of t with the energy of the signal being $\|u\|^2 = \langle u, u \rangle = \int |u(t)|^2 dt$. In solving Problem 1 we shall make use of the fact that every functional on a Hilbert space has a representation as an inner product with a fixed element of the space. Let the fixed element associated with the functional be denoted by f_1 . Then for the operator F^t the set $\{\langle f_1 \rangle_1^n\}$ in H^* has a corresponding set $\{f_1\}_1^n$ in H . It is apparent that the linear independence of the functionals implies the linear independence of the vectors.

Theorem (A): Let M denote the linear manifold spanned by the set $\{f_1\}_1^n$; then $M^\perp$, the orthogonal complement of M in U , is the null space of F^t and the restriction of F^t to M is non-singular.

Proof: Let N_{F^t} denote the null space of F^t . If $u \in M^\perp$ then $\langle f_1, u \rangle = 0, i = 1, \dots, n$ and hence $F^t u = 0$ which implies that $u \in N_{F^t}$ hence $M^\perp \subset N_{F^t}$. On the other hand if $u \in N_{F^t}$ then $F^t u = \sum_{i=1}^n \phi_i \langle f_1, u \rangle = 0$ and since the set $\{\phi_i\}_1^n$ is linearly independent

$\langle f_1, u \rangle = 0 (i = 1, \dots, n)$ which implies that $u \in M^\perp$ hence $M^\perp \supset N_{F^t}$

and the first part of Theorem 1 is proved. Let F_M denote the restriction of F^t to the subspace M that F_M is nonsingular¹ is immediate from the fact that F_M is a linear mapping between

¹If T is a linear transformation from finite dimensional space V into W , and $\dim V = \dim W$, then T is nonsingular. (33)

finite-dimensional spaces with a domain and range of equal dimensions. (33)

Theorem (B): The set U_x has a unique element in M which is exactly u° , the solution of Problem I.

Proof: Since F_M is nonsingular F_M^{-1} exists and the element defined by $u = F_M^{-1}x$ is the unique preimage of x in M . Since M is finite dimensional it is closed and we may consider the orthogonal decomposition $U = M \oplus M^\perp$ of U . Let $\bar{u}$ denote an arbitrary element of U which may then be uniquely expressed as

$$\bar{u} = \bar{u}_1 + \bar{u}_2$$

where $\bar{u}_1 \in M$ and $\bar{u}_2 \in M^\perp$. Since $M^\perp$ is the null space of F^t the additional relation

$$x = F^t \bar{u} = F^t (\bar{u}_1 + \bar{u}_2) = F^t \bar{u}_1 + F^t \bar{u}_2 = F^t \bar{u}_1 + 0 \text{ holds}$$

since $\bar{u}_1$ is a preimage of x in M it is unique and given by

$$u_1 = F_M^{-1} x$$

From the relation

$$\|\bar{u}\|^2 = \left\| F_M^{-1} x \right\|^2 + \|\bar{u}_2\|^2$$

it is apparent that

$$\|\bar{u}\|^2 \geq \left\| F_M^{-1} x \right\|^2$$

holds for every other $\bar{u} \in U_x$ with equality if and only if $\bar{u} = F_M^{-1}x$ and hence the theorem is proved.

Remarks: The transformation F_M^{-1} is exactly the pseudoinverse of F . Since M is uniquely determined by F no confusion can

result from introducing the notation $F^\dagger = F_M^{-1}$ which will be used in the sequel.

Theorem (B) satisfies the theoretical requirements of Problem 1 and it is easy to fill in the computational details. To illustrate let $x \in R^n$ be a given arbitrary vector and take the set $\{f_i\}_1^n$ as a basis for M.

Since

$$u^\circ = \sum_{j=1}^n \alpha_j f_j \quad (2.2.1-2)$$

in terms of this basis here $\alpha = (\alpha_1 \dots \alpha_n)$ and is a scalar tuplet to be determined. Similarly x has a unique expansion along the basis $\{\phi_i\}_1^n$ for R^n and

$$x = \sum_{i=1}^n \beta_i \phi_i \quad (2.2.1-3)$$

where the tuplet $\beta = (\beta_1 \dots \beta_n)$ is uniquely determined by the vector x and the basis set $\{\phi_i\}_1^n$.

Using equation 2.2.1-1 and 2.2.1-2 it is immediate that

$$x = \sum_{i=1}^n \beta_i \phi_i = F u^\circ = \sum_{i=1}^n \phi_i \left[\sum_{j=1}^n \langle f_i, f_j \rangle \alpha_j \right] \quad (2.2.1-4)$$

Since the set $\{\phi_i\}_1^n$ is linearly independent equation 2.2.1-4 implies the n equations

$$\beta_i = \sum_{j=1}^n \langle f_i, f_j \rangle \alpha_j \quad i = 1 \dots n$$

or in matrix form:

$$\beta = \mathcal{F} \alpha \quad (2.2.1-5)$$

Where f_{ij} , the (i,j) element of the matrix $\mathcal{F}$, is defined by $f_{ij} = \langle f_i, f_j \rangle$.

$$\mathcal{F} = [\langle f_1, f_j \rangle] \quad (2.2.1-6)$$

The determinant of $\mathcal{F}$ is the Gram determinant of the set $\{f_1\}_1^n$ and hence the linear independence of this set assures the nonsingularity of $\mathcal{F}$. Since the tuplet $\beta = (\beta_1 \dots \beta_n)$ is uniquely determined by x , the equation

$$\alpha = \mathcal{F}^{-1}\beta$$

uniquely determines α and in turn, through equation (2.2.1-2), u° . The matrix $\mathcal{F}$ also yields a simple expression for the energy of u° . Indeed the expression

$$\|u^\circ\|^2 = \langle u^\circ, u^\circ \rangle = \left\langle \sum_{i=1}^n \alpha_i f_i, \sum_{j=1}^n \alpha_j f_j \right\rangle = \sum_{i,j=1}^n \alpha_i \langle f_i, f_j \rangle \alpha_j$$

is precisely the quadratic form

$$\|u^\circ\|^2 = [\alpha, \mathcal{F}\alpha] = [\beta, \mathcal{F}^{-1}\beta] \quad (2.2.1-7)$$

The matrix $\mathcal{F}^{-1}$ has a similar form to that given in (2.2.1-6). Letting $\{f_1^+\}_1^n$ denote the dual basis for M it is easily verified that (see Appendix AVIII)

$$\mathcal{F}^{-1} = [\langle f_1^+, f_j^+ \rangle] \quad (2.2.1-8)$$

Remarks: The use of the $\mathcal{F}$ matrix is not the only computational method for determining u° . Let the vectors $\{\phi_1^+\}_1^n, \{f_1^+\}_1^n$ denote the dual basis for R^n and M respectively. It is apparent from equation (2.2.1-1) that the transformation $F^\dagger$ is given by

$$F^\dagger = \sum_{i=1}^n f_1^+ \langle \phi_1^+, \cdot \rangle \quad (2.2.1-9)$$

and hence it is immediate that u° is given by

$$u^\circ = \sum_{i=1}^n f_1^+ \langle \phi_1^+, x \rangle \quad (2.2.1-10)$$

The energy of u^0 is also easily computed in this case and is given by

$$\begin{aligned} \|u^0\|^2 &= \left\langle \sum_{i=1}^n f_i^+ \right\rangle \left\langle \phi_1^+, x \right\rangle, \sum_{j=1}^n f_j^+ \left\langle \phi_j^+, x \right\rangle \\ &= \sum_{i,j} \left\langle \phi_1^+, x \right\rangle \left\langle f_1^+, f_j^+ \right\rangle \left\langle \phi_j^+, x \right\rangle \quad (2.2.1-11) \end{aligned}$$

In the light of the fact that $\beta_1 = \left\langle \phi_1^+, x \right\rangle$ and the preceding remark it is obvious that this quadratic form is exact equation (2.2.1-7). Before proceeding to the theoretical generalization of the present discussion, it is fruitful to consider two concrete examples from the theory of linear dynamic systems.

2.2.2 Examples

Example 1: Continuous Linear System

For the real valued linear dynamic system defined by

$$\dot{x}(t) = A(t) x(t) + B(t) u(t) \quad (2.2.2-1)$$

let us determine the control $u(t)$ which transfers the vector $x(t_0) = x_0$ to the value $x(t_f) = x_f$ while minimizing the performance index

$$J = 1/2 \int_{t_0}^{t_f} [u(s), u(s)] ds \quad (2.2.2-2)$$

To formulate this problem in the notation of the main discussion, it is only necessary to introduce the input space $U = L_2^m(t_0, t_f)$, the direct sum of m copies of the real component Hilbert space $L_2(t_0, t_f)$. The inner product on U is given by

$$\langle u, v \rangle = \int_{t_0}^{t_f} [u(s), v(s)] ds \quad u, v \in U \quad (2.2.2-3)$$

and thus $J(s) = 1/2 \|u\|^2$. (2.2.2-4)

This example may be reformulated in the following manner:

given x_0 and x_f in R^n the vector $\hat{x}(t_f) \in R^n$ is defined by $\hat{x}(t_f) = [x(t_f) - \phi(t_f, t_0)x_0]$ and the minimizing control is the element u with the minimum norm such that $\hat{x} = F^t u$. This is exactly Problem 1 discussed above.

In the present instance, the columns $\{\phi_1(t_f)\}_1^n$ of the transition matrix $\phi(t_f, t_0)$ and $\{f_1\}_1^n$ serve as bases for R^n and M respectively. It is not difficult to show that the solution u^0 is given by the expression

$$u^0(t) = \sum_{i=1}^n f_i \alpha_i = B^*(t) \phi^*(t_0, t) \mathcal{F}^{-1}(t_f) \beta \quad (2.2.2-5)$$

where f_i = i -th row of matrix $\phi(t_0, t)B(t)$ (See Appendix AIII)

$$\beta = \phi(t_0, t_f) \hat{x}(t_f) = \phi(t_0, t_f) [x_f - \phi(t_f, t_0)x_0] \quad (2.2.2-6)$$

and

$$\mathcal{F}(t_f) = \int_{t_0}^{t_f} \phi(t_0, \tau) B(\tau) B^*(\tau) \phi^*(t_0, \tau) d\tau \quad (2.2.2-7)$$

Example 2: Discrete Linear System

Consider the real valued discrete dynamic system defined by

$$\Delta_t x(t) = A(t) x(t) + B(t) u(t), \quad t \in \delta \quad (2.2.2-8)$$

The problem posed is to find the control u defined on the set $\delta = \{t_0, t_1 \dots t_f\}$ such that transfers the tuplet (x_0, t_0) to the tuplet (x_f, t_f) while minimizing the

performance index

$$J = \sum_{j=0}^{[t_f]-1} h_j [u(t_j), u(t_j)] \quad (2.2.2-9)$$

The similarity with the previous example makes the present case easy. If $\ell_2(\delta)$ denotes the Hilbert function space of all real functions defined on δ with the inner product

$$\langle u, v \rangle = \sum_{\alpha} h_{\alpha} u(t_{\alpha}) v(t_{\alpha}) \quad u, v \in \ell_2(\delta) \quad (2.2.2-10)$$

then taking the input space $U = \ell_2^m(\delta)$ with the inner product

$$\langle u, v \rangle = \sum_{t_j \in \delta} h_j [u(t_j), v(t_j)] \quad u, v \in U$$

it is clear that the present problem is closely analogous to Example 1. It is easily verified that the matrix $\mathcal{F}$ is given by

$$\mathcal{F}(t_f, t_0) = \sum_{\alpha=[t_0]}^{[t_f]-1} h_{\alpha} \Phi(t_0, t_{\alpha+1}) B(t_{\alpha}) B^*(t_{\alpha}) \Phi^*(t_0, t_{\alpha+1}) \quad (2.2.2-11)$$

while the minimizing function u^0 is computed by the rule

$$u^0(t_j) = B^*(t_j) \Phi^*(t_0, t_{j+1}) \mathcal{F}^{-1}(t_f, t_0) [\Phi(t_0, t_f) x_f - x_0] \quad (2.2.2-12)$$

2.2.3 Physical Applications

Example 1: Consider the continuous second order position control servomechanism shown in Figure 2.2.3-1 in which the gain, mass, and viscous damping which in general are deterministic differentiable functions of time are assumed in this example to be constant. The motor armature inductance and backlash of the gear are neglected. Now by fixing t_f the following basic equations describe this system

$$e_1 = A e_r = A k_4 (\theta_1 - \theta_p) = R\dot{\theta}_o + k_1 \dot{\theta}_o \quad (2.2.3-1)$$

$$T = k_2 i = J \ddot{\theta}_o + f \dot{\theta}_o \quad (2.2.3-2)$$

$$\theta_p = k_3 \theta_o \quad (2.2.3-3)$$

where A = gain of amplifier

k_1 = motor constant

T, k_2 = developed torque, torque constant

k_3 = gear ratio

k_4 = proportionality constant of potentiometers

ϕ = air gap flux of motor

R = resistance of armature circuit of servomotor M

R_f, L_f = resistance, inductance of motor field circuit

J, f = inertia, viscous damping (referred to motor shaft)

It is desired to find $\theta_1(t)$ such that

$$\theta_o(t_o) = \theta_o, \quad \dot{\theta}_o(t_o) = \dot{\theta}_o, \quad \theta_o(t_f) = \theta_f \text{ and } \dot{\theta}_o(t_f) = \dot{\theta}_f$$

and

$$J = 1/2 \int_{t_o}^{t_f} \dot{\theta}_o^2 dt \text{ has a minimum value.}$$

From equations (2.2.3-1) to (2.2.3-3) we have

$$\frac{JR \ddot{\theta}_o}{k_2} + \left(\frac{fR}{k_2} + k_1 \right) \dot{\theta}_o + A k_4 k_3 \theta_o = A k_4 \theta_1 \quad (2.2.3-4)$$

or
$$\ddot{\theta}_o + a_1 \dot{\theta}_o + a_o \theta_o = d \theta_1$$

where

$$a_1 = \frac{k_2}{JR} \left(\frac{fR}{k_2} + k_1 \right)$$

$$a_0 = \frac{A k_2 k_3 k_4}{JR}$$

$$d = A k_2 k_4 \left(\frac{1}{JR} \right)$$

Let $x_1 = \theta_o$, $x_2 = \dot{\theta}_o$ and $u = \theta_1$

Then in matrix notation equation (2.2.3-4) may be written

$$\dot{x}(t) = A x(t) + B u(t) \quad (2.2.3-5)$$

where $A = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 \\ d \end{bmatrix}$

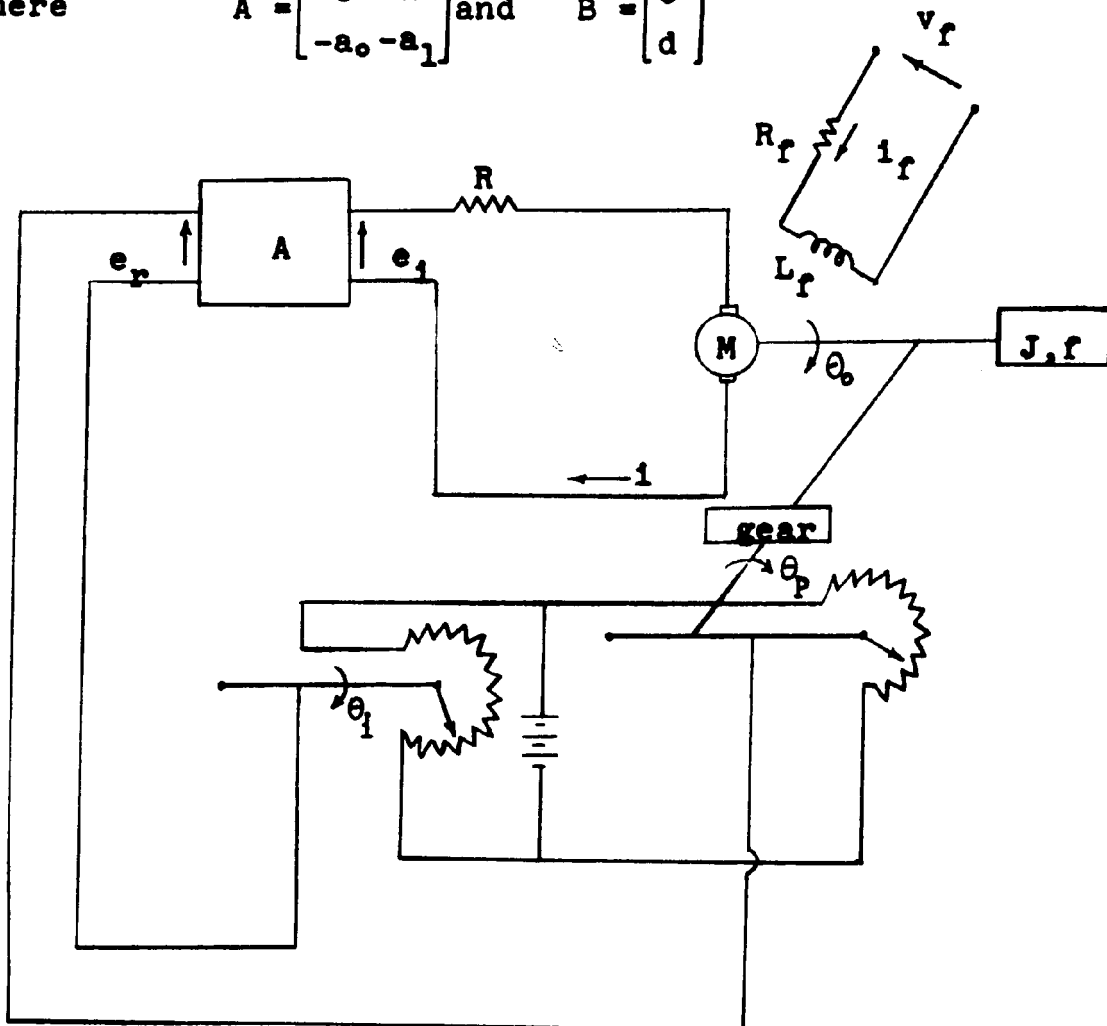


Fig. 2.2.3-1.--Second order position control servomechanism.

It is not difficult to show that the transition matrix $\Phi(t, t_0)$ of the system (2.2.3-5) is given by

$$\Phi(t, t_0) = \begin{bmatrix} \phi_{11}(t, t_0) & \phi_{12}(t, t_0) \\ \phi_{21}(t, t_0) & \phi_{22}(t, t_0) \end{bmatrix}$$

where

$$\phi_{11}(t, t_0) = c_1 e^{-b(t-t_0)} \sin[w(t-t_0) + \psi_1]$$

$$\phi_{12}(t, t_0) = c_2 e^{-b(t-t_0)} \sin w(t-t_0)$$

$$\phi_{21}(t, t_0) = c_3 e^{-b(t-t_0)} \sin w(t-t_0)$$

$$\phi_{22}(t, t_0) = c_4 e^{-b(t-t_0)} \sin[w(t-t_0) + \psi_2]$$

$$c_1 = \frac{(b^2 + w^2)^{1/2}}{w}, \quad c_2 = -\frac{1}{w}, \quad c_3 = \frac{a_0}{w}, \quad c_4 = \frac{[(a_1 - b)^2 + w^2]^{1/2}}{w}$$

$$w = (a_0 - b^2)^{1/2}, \quad b = \frac{a_1}{2}, \quad \psi_1 = \tan^{-1}\left(\frac{w}{-b}\right), \quad \text{and} \quad \psi_2 = \tan^{-1}\left(\frac{w}{a_1 - b}\right)$$

The initial and final value of the state of the system can be expressed by

$$x(t_0) = \begin{bmatrix} \theta_0(t_0) \\ \dot{\theta}_0(t_0) \end{bmatrix} = \begin{bmatrix} \theta_0 \\ \dot{\theta}_0 \end{bmatrix}$$

$$x(t_f) = \begin{bmatrix} \theta_0(t_f) \\ \dot{\theta}_0(t_f) \end{bmatrix} = \begin{bmatrix} \theta_f \\ \dot{\theta}_f \end{bmatrix}$$

respectively. We can compute

$$\begin{aligned}
\beta &= \Phi(t_0, t_f) x(t_f) - x(t_0) \\
&= \begin{bmatrix} \phi_{11}(t_0, t_f) & \phi_{12}(t_0, t_f) \\ \phi_{21}(t_0, t_f) & \phi_{22}(t_0, t_f) \end{bmatrix} \begin{bmatrix} \theta_f \\ \dot{\theta}_f \end{bmatrix} + \begin{bmatrix} \theta_0 \\ \dot{\theta}_0 \end{bmatrix} \\
&= \begin{bmatrix} \phi_{11}(t_0, t_f)\theta_f + \phi_{12}(t_0, t_f)\dot{\theta}_f + \theta_0 \\ \phi_{21}(t_0, t_f)\theta_f + \phi_{22}(t_0, t_f)\dot{\theta}_f + \dot{\theta}_0 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}
\end{aligned}$$

where

$$d_1 = \phi_{11}(t_0, t_f)\theta_f + \phi_{12}(t_0, t_f)\dot{\theta}_f + \theta_0$$

$$d_2 = \phi_{21}(t_0, t_f)\theta_f + \phi_{22}(t_0, t_f)\dot{\theta}_f + \dot{\theta}_0$$

$$\begin{aligned}
&= \int_{t_0}^{t_f} \Phi(t_0, s) B(s) B^*(s) \Phi^*(t_0, s) ds \\
&= \int_{t_0}^{t_f} d^2 \begin{bmatrix} \phi_{12}^2(t_0, s) & \phi_{12}(t_0, s) \phi_{22}(t_0, s) \\ \phi_{22}(t_0, s) \phi_{12}(t_0, s) & \phi_{22}^2(t_0, s) \end{bmatrix} ds \\
&= \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}
\end{aligned}$$

where

$$f_{11} = \int_{t_0}^{t_f} d^2 \phi_{12}^2(t_0, s) ds$$

$$f_{12} = \int_{t_0}^{t_f} d^2 \phi_{12}(t_0, s) \phi_{22}(t_0, s) ds$$

$$f_{21} = \int_{t_0}^{t_f} d^2 \phi_{22}(t_0, s) \phi_{12}(t_0, s) ds$$

$$f_{22} = \int_{t_0}^{t_f} d^2 \phi_{22}^2(t_0, s) ds$$

$$\mathcal{F}^{-1} = \frac{1}{(f_{11}f_{22} - f_{12}f_{21})} \begin{bmatrix} f_{22} & -f_{12} \\ -f_{21} & f_{11} \end{bmatrix}$$

Then the optimal control $u^\circ(t)$ or $\theta_i^\circ(t)$ is given by

$$u^\circ(t) = \sum_{i=1}^2 f_i \alpha_i$$

$$= B^*(t) \Phi^*(t_0, t) \mathcal{F}^{-1} \beta$$

$$= [0, d] \begin{bmatrix} \phi_{11}(t_0, t) & \phi_{21}(t_0, t) \\ \phi_{12}(t_0, t) & \phi_{22}(t_0, t) \end{bmatrix} \frac{1}{(f_{11}f_{22} - f_{12}f_{21})} \begin{bmatrix} f_{22} & -f_{12} \\ -f_{21} & f_{11} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

or

$$\theta^\circ(t) = u^\circ(t) = \frac{d}{(f_{11}f_{22} - f_{12}f_{21})} (d_1 f_{22} \phi_{12} - d_2 f_{12} \phi_{12} - d_1 f_{21} \phi_{22} + d_2 f_{11} \phi_{22})$$

and the optimal performance index can be computed as follows:

$$J^\circ = \langle \beta, \mathcal{F}^{-1} \beta \rangle$$

$$= \frac{1}{(f_{11}f_{22} - f_{12}f_{21})} [d_1, d_2] \begin{bmatrix} f_{22} & -f_{12} \\ -f_{21} & f_{11} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

$$= \frac{1}{(f_{11}f_{22} - f_{12}f_{21})} (d_1^2 f_{22} - d_1 d_2 f_{12} - d_1 d_2 f_{21} + d_2^2 f_{11})$$

Given specific values of these constants J , f , R , A , k_1 , k_2 , k_3 and k_4 we can compute $\theta^\circ(t)$ and J° . This method is applicable to high order systems, however the size of the related transition matrix increases.

Example 2: Consider the sampled-data control system given by Figure 2.2.3-2:

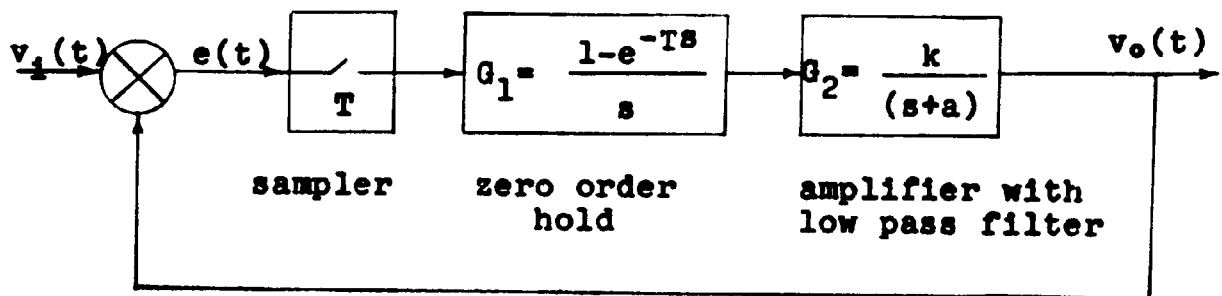


Fig. 2.2.3-2.--First order system with sampler and zero-order hold.

Assuming for simplicity that a , k , and T are all unity, one can easily verify (82) that

$$\frac{v_o(z)}{v_1(z)} = \frac{G(z)}{1 + HG(z)}$$

where

$$G(s) = \frac{1 - e^{-s}}{s} \times \frac{1}{s+1}$$

$$G(z) = \frac{z(1-e^{-1})}{(z-1)(z-e^{-1})} - \frac{1-e^{-1}}{(z-1)(z-e^{-1})} = \frac{1-e^{-1}}{z - e^{-1}}$$

After some manipulation we have

$$\frac{v_o(z)}{v_1(z)} = \frac{0.632}{z + 0.264} \quad (2.2.3-6)$$

Let $v_0 = x$, and $v_1 = u$, equation (2.2.3-6) reduces to

$$zx(z) + 0.264 x(z) = 0.632u(z)$$

which leads to the following first order linear difference equation

$$x(k+1) + 0.264 x(k) = 0.632u(k) \quad (2.2.3-7)$$

$$\text{or} \quad \Delta x(k) = -1.264 x(k) + .632u(k) \quad (2.2.3-8)$$

$$\text{where} \quad \Delta x(k) = x(k+1) - x(k)$$

Replacing k by t where $t \in \mathcal{S} = [t_0, t_1, \dots, t_f]$ equation (2.2.3-8) may be written in the following form

$$\Delta x(t) = -1.264 x(t) + 0.632u(t) \quad (2.2.3-9)$$

The transition matrix of system (2.2.3-9) is given by

$$\Phi(t, t_0) = (-.264)^{t-t_0} \text{ which satisfies}$$

$$\Delta \Phi(t, t_0) = 1.264 \Phi(t, t_0) \quad (2.2.3-10)$$

The optimal input function $u(t)$ that transfers the output from $x(t_0) = 0$ to $x(t_f) = x_f$ while minimizing the performance index

$$J = \sum_{j=0}^{[t_f]-1} h_j [u(t_j), u(t_j)] = \sum_{j=0}^{[t_f]-1} u^2(t_j),$$

$$(\text{where } h_j = t_{k+1} - t_k = 1)$$

is given by (see equation 2.2.2-12)

$$\begin{aligned} u(t_j) &= B^*(t_j) \Phi^*(t_0, t_{j+1}) \mathcal{F}^{-1}(t_f, t_0) [\Phi(t_0, t_f) x_f - x_0] \\ &= .632(-.264)^{(t_0-t_{j+1})} \left[\sum_{\alpha=t_0}^{[t_f]-1} 4(-.264)^{2(t_0-t_{\alpha+1})} \right]^{-1} (-.264)^{t_0-t_f} x_f \\ &= .632(-.264)^{(2t_0-t_{j+1}-t_f)} \left[\sum_{\alpha=[t_0]}^{[t_f]-1} 4(-.264)^{2(t_0-t_{\alpha+1})} \right]^{-1} x_f \end{aligned}$$

where

$$\begin{aligned} \mathcal{F}(t_f, t_0) &= \sum_{\alpha=[t_0]}^{[t_f]-1} h_{\alpha} \tilde{\Phi}(t_0, t_{\alpha+1}) B(t_{\alpha}) B^{\#}(t_{\alpha}) \tilde{\Phi}^{\#}(t_0, t_{\alpha+1}) \\ &= \sum_{\alpha=[t_0]}^{[t_f]-1} 4 (-0.264)^{2(t_0 - t_{\alpha+1})} \end{aligned}$$

A simple extension to Problem 1 is given by theorem

C. Let u_0 be an arbitrary element of H and $x_0 = Fu_0$ its image under F .

Theorem C: v = the element $v \in U_x$ which minimizes

$$\begin{aligned} J(u) &= \|u_0 + v\|^2 \text{ is given by} \\ F^{\dagger}(x + x_0) - u_0 \end{aligned}$$

Proof:

Let v be defined as $F^{\dagger}(x + x_0) - u_0$; then

$$Fv = F\{F^{\dagger}(x + x_0) - u_0\} = FF^{\dagger}(x + x_0) - Fu_0 = x + x_0 - x_0 = x$$

and hence $v \in U_x$. The equation $v + u_0 = F^{\dagger}(x + x_0)$ implies, however, that $v + u_0$ is the element of minimum norm which maps onto $x + x_0$ under F . Since the norm of $v + u_0$ is J the theorem is proved. From the solution given above to Problem 1 it is clear that v can be computed by

$$v + u_0 = \sum_{i=1}^n f_i \alpha_i$$

$$\text{where } \alpha = \mathcal{F}^{-1} \beta$$

Example: Since the present discussion is so closely linked with Problem 1 above, this example is selected not as a computational illustration but rather as a suggestion of a

class of physical problems which give rise to the problem posed in theorem C.

Consider a non-linear discrete dynamic system defined by

$$\Delta_t x(t) = f(x(t), u(t), t) \quad t \in \delta$$

Let the tuplet $(\bar{x}, \bar{u})$ denote a trajectory for this system which is optimal for an arbitrary performance index. As the control $\bar{u}$ is applied we shall assume that slight inaccuracies in the control u , or inherent limitations of the knowledge of the vector function $f(x, u, t)$ cause the actual trajectory to deviate slightly from $\bar{x}$. Letting δx and δu denote deviations from $(\bar{x}, \bar{u})$ and assuming that f is continuous with respect to $(\bar{x}, \bar{u})$ it is not difficult to verify that the perturbation tuplet $(\delta x, \delta u)$ satisfies

$$\Delta_t \delta x(t) = A(t) \delta x(t) + B(t) \delta u(t)$$

where the (i, j) entry of $A(t)$ is given by

$$a_{ij}(t) = \left. \frac{\partial f_i}{\partial x_j} \right|_{(\bar{x}, \bar{u})} \quad i, j = 1..n$$

and the typical entry of $B(t)$ is exactly

$$b_{ij}(t) = \left. \frac{\partial f_i}{\partial u_j} \right|_{(\bar{x}, \bar{u})} \quad i = 1..n, j = 1..m.$$

As a practical measure δu is to be chosen to produce a desired value $\delta x(t_f)$. While the original criterion of optimality may have been something quite different from energy, the fact that the physical mission may require that the fuel available is actually the pre-computed value $\|\bar{u}\|^2$ makes a minimum energy strategy for the perturbation control essential.

Here we may choose to minimize $\|\delta u\|^2$ as in Problem 1, however since the actual total energy being used is given by $\|\bar{u} + \delta u\|^2$ it might be more appropriate to minimize this quantity which brings the problem within the scope of theorem C.

The class of problems with which we have been concerned in the previous paragraphs was centered about a transformation of a Hilbert function space U onto R^n . In the succeeding discussions the scope of interest is given a natural generalization patterned after the behavior of linear dynamic systems. While the details become more complex, the net result is that in spite of the complexity the general problems to be considered are exactly the same as Problem 1.

The solution of Problem 1 is the basic result on the minimum energy problem. In this and subsequent sections this basic result will be expanded in scope by consideration of several variations on the basic problem. The first such consideration is a variable end point problem which will be studied in the following section.

2.3 Problem 2

Consider the parametrized family $\{F^s : s \in \mathcal{V}\}$ of linear transformations each of which is a mapping of the Hilbert function space U onto R^n . Thus for each $s \in \mathcal{V}$ the operator F^s of this family has the form

$$F^s = \sum_{i=1}^n \phi_i(s) \langle f_i^s, \cdot \rangle \quad (2.3-1)$$

where the vectors $\{\phi_1(s)\}_1^n$ and the functionals $\{\langle f_1^s \rangle_1^n$ are linearly independent sets in R^n and U^* respectively. Find the vector $x \in R^n$, the parameter $s \in \mathcal{V}$ and the input $u \in U$ such that the constraints

$$x = F^s u \quad s \in \mathcal{V} \quad (2.3-2)$$

$$\psi_1(x, s) = 0 \quad i = 1 \dots m < n, \quad x \in R^n, \quad s \in \mathcal{V} \quad (2.3-3)$$

are satisfied and the function

$$J = g(x, s) + \|u\|^2 \quad (2.3-4)$$

is minimized.

2.3.1 Abstract Theory

In solving this problem it is desirable to take advantage of the results of Problem 1. To do this let us note that in the family $\{F^s : s \in \mathcal{V}\}$ each of M^s is the orthogonal complement of the null space of the respective transformation F^s . The restriction of F^s to M^s is a non-singular mapping onto R^n hence an associated family of pseudoinverse operators $\{(F^s)^\dagger : s \in \mathcal{V}\}$ may be defined such $(F^s)^\dagger F^s$ is the orthogonal projection of U on M^s for all $s \in \mathcal{V}$.

Assume now that Problem 2 has a solution and denote by $(\bar{x}, \bar{s})$ the range vector and the parameter value which partially solves Problem 2. Let us assume for the moment that $(\bar{x}, \bar{s})$ is known, and after substituting $(\bar{x}, \bar{s})$ into (2.3-3) and (2.3-4) the functions $g(x, s)$ and $\psi_1(x, s)$ $i = 1 \dots m$ are independent of u , and the Problem 2 is reduced

to Problem 1 with fixed $(\bar{x}, \bar{s})$. To minimize J it is clear that $\bar{u}$ the solution to Problem 2 is chosen as the solution to Problem 1. Hence $\bar{u} = (F^{\bar{s}})^{\dagger} \bar{x}$.

Remarks: These remarks show that Problem 2 may be stated in the following equivalent form: Let the transformation $(F^s)^{\dagger}$ defined by

$$(F^s)^{\dagger} \triangleq \sum_{j=1}^n (r^s)_j^+ > < \phi_1^+ (s)$$

denote a typical member of the set $\{(F^s)^{\dagger} : s \in \mathcal{V}\}$ and let the set X be defined by

$$X \triangleq \{(x, s) : x \in \mathbb{R}^n, s \in \mathcal{V}, \psi_1(x, s) = 0, i = 1 \dots m\}$$

The solution of Problem 2 is given by the tuplet

$(\bar{x}, \bar{s}, (F^{\bar{s}})^{\dagger} \bar{x})$ for which $(\bar{x}, \bar{s})$ satisfies

$$J(\bar{x}, \bar{s}) = \inf_{(\bar{x}, \bar{s}) \in X} \left\{ g(\bar{x}, \bar{s}) + \left\| \sum_{i=1}^n (r^{\bar{s}})_i^+ > < \phi_1^+ (\bar{s}), \bar{x} \right\|^2 \right\} \quad (2.3.1-1)$$

This formulation may also be expressed in matrix form. Let

$$\beta_1(s) = < \phi_1^+ (s), x > \quad i=1 \dots n \text{ and } \beta(s) = (\beta_1(s) \dots \beta_n(s)).$$

It is clear that the functions x_i and g may be expressed as functions of β_1 thus the set X becomes

$$X = \{(\beta(s), s) : s \in \mathcal{V}, \psi_1(\beta(s), s) = 0, i = 1 \dots m\}$$

and the solution to Problem 2 is determined by $(\beta(\bar{s}), \bar{s})$

such that

$$J(\beta(\bar{s}), \bar{s}) = \inf_{(\beta(\bar{s}), \bar{s}) \in X} \left\{ g(\beta(\bar{s}), \bar{s}) + [\beta(\bar{s}), \mathcal{F}^{-1}(\bar{s}) \beta(\bar{s})] \right\} \quad (2.3.1-2)$$

Evidently this is an ordinary constrained minimization problem.

It is not our intention to explore in depth the many variations and examples which stem from equation (2.3.1-2). When all functions involved are sufficiently smooth, the classical statement of necessary and sufficient conditions is given by the theorem.

Theorem D: Let $f(x_1 \dots x_n)$ denote an objective function to be extremalized which in addition to the r independent side constraints

$$g_j(x_1 \dots x_n) = 0 \quad j=1 \dots r < n \quad (2.3.1-3)$$

are at least twice differentiable at a point $(\bar{x}_1 \dots \bar{x}_n)$.

If $\bar{x} = (\bar{x}_1 \dots \bar{x}_n)$ is a local constrained extremum for $f(x)$ then it is necessary that scalars $(\lambda_1 \dots \lambda_r)$ exist for which

$$J^* = f(x_1 \dots x_n) + \sum_{j=1}^r \lambda_j g_j(x_1 \dots x_n) \quad (2.3.1-4)$$

has an unconstrained extremum at $\bar{x}$. If in addition the relationship

$$\sum_{i,j=1}^n \frac{\partial^2 J^*}{\partial x_i \partial x_j} \bigg|_{\bar{x}} z_i z_j \geq 0 \quad (2.3.1-5)$$

holds for all $z = (z_1 \dots z_n)$ satisfying

$$\sum_{i=1}^n \frac{\partial g_j}{\partial x_i} \bigg|_{\bar{x}} z_i = 0 \quad (j=1, \dots, r) \quad (2.3.1-6)$$

then $\bar{x}$ yields a constrained minimum of f .

Remarks: (i) If $\bar{x}$ is an unconstrained extremum of J^* ($\bar{x}$) then the $n+r$ equations

$$\frac{\partial J^*(\bar{x})}{\partial x_1} = 0 \quad i = 1 \dots n \quad (2.3.1-7)$$

$$g_j(\bar{x}) = 0 \quad j = 1 \dots r \quad (2.3.1-8)$$

in the $n+r$ unknowns $(x_1 \dots x_n, \lambda_1 \dots \lambda_r)$ must be satisfied.

(ii) If H is the matrix with (i,j) element $\frac{\partial^2 f(\bar{x})}{\partial x_i \partial x_j}$ and

$\nabla_x g$ denotes the tuple $\nabla_x g = (\frac{\partial g}{\partial x_1} \dots \frac{\partial g}{\partial x_n})$ then the sufficient conditions may be phrased as being

$$[z, Hz] \geq 0 \quad (2.3.1-9)$$

for all $z = z_1 \dots z_n$ satisfying

$$[\nabla_x g_j, z] = 0 \quad (j=1, \dots, r) \quad (2.3.1-10)$$

A proof of theorem D along with geometric interpretations of Lagrange multipliers has been given by Kao (40). Other sources on this topic include (13) and (28).

2.3.2 Example: Continuous linear system with movable boundary condition

Here we shall consider again the continuous system of Example 1 in section 2.2.2. In the present instance the terminal conditions $x(t_f) = 0$ and $\|x(t_0)\|^2 = 1$ must be satisfied while minimizing the performance index $J = \|u(t)\|^2$. In short (x_0, x_f) must satisfy the constraints

$$\begin{aligned} \psi_1(x_0) &= [x_0, x_0] - 1 = 0 \\ \psi_2(x_f) &= x_f = 0 \end{aligned} \quad (2.3.2-1)$$

Drawing upon previous discussions it is clear that

$$\begin{aligned} x &= x(t_f) - \Phi(t_f, t_0)x_0 \\ &= 0 - \Phi(t_f, t_0)x_0 \end{aligned} \quad (2.3.2-2)$$

As before the scalar tuplet β will be defined by $\beta = \Phi(t_0, t_f)\hat{x}$ and in light of equation 2.3.2-2 $\beta = -x^0$; thus the first constraint may be re-written

$$\psi_1(\beta) = [\beta, \beta] - 1 = 0.$$

In this example the function g of equation (2.3-4) is identically zero, thus the performance index is written simply as

$$J(\beta) = [\beta, \mathcal{F}^{-1}\beta]$$

To minimize $J(\beta)$ while satisfying $\psi_1(\beta) = 0$ the augmented functional

$$J^*(\beta) = [\beta, \mathcal{F}^{-1}\beta] + \lambda \{ [\beta, \beta] - 1 \}$$

is introduced. The first necessary condition is given by the equations

$$\nabla_{\beta} J^*(\beta) = 2 \mathcal{F}^{-1}\beta + 2\lambda\beta = 0$$

$$\frac{\partial J^*}{\partial \lambda}(\beta) = [\beta, \beta] - 1 = 0$$

Taking the inner product with β on both sides of the first expression and utilizing the second yields the condition

$$\lambda = -[\beta, \mathcal{F}^{-1}\beta] \quad (2.3.2-3)$$

hence the defining relationship for β becomes

$$\mathcal{F}^{-1}\beta = J(\beta)\beta \quad (2.3.2-4)$$

In other words β is an eigenvector of the matrix $\mathcal{F}^{-1}$.

Since this matrix is symmetric and positive definite all eigenvalues of $\mathcal{F}^{-1}$ are real and greater than zero. From

equation (2.3.2-4) it is clear that the eigenvalues of $\mathcal{F}^{-1}$ are the numbers $J(\beta)$ where β runs through the eigenvectors. Since $J(\beta)$ is to be minimized the solution is determined by choosing β to be any unit eigenvector corresponding to the smallest eigenvalue of $\mathcal{F}^{-1}$.

2.4 Problem 3

Let H_1 and H_2 be two Hilbert function spaces and F be a transfer operator carrying H_1 into H_2 . Let F^t be the t -th sample of F a mapping of H_1 onto R^n . If U_x is the set defined by

$$U_x = \{ u : u \in H_1, F^t u = x \}$$

find the element $u \in U_x$ for which

$$J(u) = \|Fu\|_{H_2}^2 + \|u\|_{H_1}^2$$

is minimized.

Note that $\|Fu\|_{H_2}^2$ is the norm $\langle Fu, Fu \rangle$ in the Hilbert

function space H_2 and not in R^n and $\|u\|_{H_1}^2$ is the norm in H_1 .

In solving Problem 3 we shall need to recall the definition of the graph of a transformation.

2.4.1 Abstract Theory

Definition: Let F be a linear mapping of H_1 into H_2 . The set $H(F)$ defined by $H(F) = \{ (u, v) : v = Fu, u \in H_1 \}$ is called the graph of F . The graph of a transformation is thus the subset of

$$H_1 \times H_2 = \{ (u, v) : u \in H_1, v \in H_2 \}$$

for which the ordinate v is related to the abscissa u by $v = Fu$. The basic properties pertinent to the present undertaking are summarized by the Lemma (1).

Lemma (1): (1) If addition and scalar multiplication are defined on the set $H(F)$ by

$$\alpha(f, g) = (\alpha f, \alpha g)$$

$$(f_1, g_1) + (f_2, g_2) = (f_1 + f_2, g_1 + g_2) \quad (2.4.1-1)$$

then $H(F)$ is a linear space.

(ii) If H_1 and H_2 are Hilbert spaces and an inner product on $H(F)$ is defined by

$$\langle (f_1, g_1), (f_2, g_2) \rangle_{H(F)} = \langle f_1, f_2 \rangle_{H_1} + \langle g_1, g_2 \rangle_{H_2} \quad (2.4.1-2)$$

then $H(F)$ is a Hilbert space.

(iii) If F is bounded then $H(F)$ is a closed linear subspace of $H_1 \times H_2$.

Remarks: The graph of F has a natural connection to

Problem 3 for if $\omega = (u, Fu) \in H(F)$ then clearly

$$\|\omega\|^2 = \langle \omega, \omega \rangle_{H(F)} = \langle u, u \rangle_{H_1} + \langle Fu, Fu \rangle_{H_2} = J(u) \quad (2.4.1-3)$$

In other words the graph of F is a Hilbert space for which $J(u)$ is the square of norm.

Let G^t denote the linear transformation of $H(F)$ onto R^n with the property

$$G^t(u, Fu) = F^t u = x(t) \quad u \in H_1 \quad (2.4.1-4)$$

It is clear that G^t is linear and well defined on $H(F)$.

Thus if $x(t) \in R^n$ and if $\omega \in H(F)$ is the element with minimum norm mapping into $x(t)$ under G^t then u the abscissa of ω maps

into $x(t)$ under F^t and is the $u \in H_1$ which minimizes

$$J(u) = \|Fu\|_{H_2}^2 + \|u\|_{H_1}^2$$

while doing so this Problem 3 reduces to Problem 1 with

$$G^t : H(F) \rightarrow R^n$$

being the operator in question.

The solution of Problem 1 was given by restricting the transformation F^t to the orthogonal complement of its null space. This restriction of F^t was nonsingular and its inverse maps any x in the range of F^t back into its preimage with minimum norm. This process will work as well for the present situation and we shall proceed to locate the orthogonal complement of the null space of G^t .

Lemma (2) The null space of G^t , denoted by N_G^t is given by the set

$$N_{G^t} = \left\{ (u, Fu) : u \in N_{F^t} \right\} \quad (2.4.1-5)$$

Proof: If $u \in N_{F^t}$ then $G^t (u, Fu) = F^t u = 0$ hence

$N_{F^t} \subset N_{G^t}$. Similarly if $(u, Fu) \in N_{G^t}$ then $G^t (u, Fu) = 0 = F^t u$

hence $N_{G^t} \subset N_{F^t}$; thus $N_{G^t} = N_{F^t}$.

Theorem E: Let Q and M denote the orthogonal complements of the null spaces of the transformations G^t and F^t respectively then Q is given by

$$Q = \left\{ (u, Fu) : (I + F^*F) u \in M \right\} \quad (2.4.1-6)$$

Proof: The proof of theorem E is given by the following chain of set equalities

$$\begin{aligned}
Q &= \left\{ (u, Fu) : (u, Fu) \perp (v, Fv) \text{ all } v \in N_{Ft} \right\} \\
&= \left\{ (u, Fu) : \langle u, v \rangle_{H_1} + \langle Fu, Fv \rangle_{H_2} = 0, v \in N_{Ft} \right\} \\
&= \left\{ (u, Fu) : \langle u, v \rangle_{H_1} + \langle F^* Fu, v \rangle_{H_1} = 0, v \in N_{Ft} \right\} \quad (\text{see} \\
&\hspace{15em} \text{Appendix AI-15}) \\
&= \left\{ (u, Fu) : \langle (I + F^* F) u, v \rangle_{H_1} = 0, v \in N_{Ft} \right\} \\
&= \left\{ (u, Fu) : (I + F^* F) u \in M \right\} \quad (2.4.1-7)
\end{aligned}$$

Remarks: It is occasionally convenient to deal with the projection of Q onto H_1 . This set will be denoted by S_u . It is clear from theorem E that S_u may be specified in any of the equivalent ways

$$\begin{aligned}
S_u &= \left\{ u : (u, Fu) \in Q \right\} \\
&= \left\{ u : (I + F^* F) u \in M \right\} \\
&= \left\{ u : u = (I + F^* F)^{-1} v, v \in M \right\} \\
&= \left\{ u : u \in (I + F^* F)^{-1} M \right\} \quad (2.4.1-8)
\end{aligned}$$

Here the fact that $(I + F^* F)$ is nonsingular has been used. The following lemma justifies this assumption.

Lemma (3) If F is a linear transformation of H_1 into H_2 , then $I + F^* F$ is nonsingular (1) and (79).

Proof: We must show that $I + F^* F$ is one to one onto as a mapping of H_1 into itself.

(1) It is one to one, because the equation $(I + F^* F)u = 0$, $u \in H_1$ implies that

$$\begin{aligned}
0 &= \langle (I + F^*F)u, u \rangle \\
&= \langle u, u \rangle_{H_1} + \langle F^*Fu, u \rangle_{H_2} \\
&= \langle u, u \rangle_{H_1} + \langle Fu, Fu \rangle_{H_2} \\
&= \|u\|_{H_1}^2 + \|Fu\|_{H_2}^2
\end{aligned}$$

from which we have $u = 0$.

(11) We next show that the range R of $I + F^*F$ is closed.

It follows from

$$\|(I + F^*F)u\|_{H_1}^2 = \|u\|_{H_1}^2 + \|F^*Fu\|_{H_1}^2 + 2\langle F^*Fu, u \rangle_{H_1} \quad (2.4.1-9)$$

that $\|(I + F^*F)u\|_{H_1} \geq \|u\|_{H_1}$

since F^*F is positive. By this inequality and the completeness of H_1 , R is complete and therefore closed. We conclude the proof that $R = H$; for otherwise there would exist a nonzero vector u_0 orthogonal to R and this contradicts the fact that

$$\langle (I + F^*F)u_0, u_0 \rangle = 0 \Rightarrow \|u_0\|_{H_1}^2 = -\langle F^*Fu_0, u_0 \rangle_{H_1} < 0 \Rightarrow u_0 = 0.$$

The solution to Problem 3 can now be concisely formulated. Indeed the minimizing element u for the functional $J(u) = \|Fu\|_{H_2}^2 + \|u\|_{H_1}^2$

while satisfying $x = F^t u$ is the unique element in S_u which is a preimage of x under F^t . It is also determined by the condition that it is the image under $(I + F^*F)^{-1}$ of the

unique element $v \in M$ for which

$$x = F^t (I + F^*F)^{-1}v \text{ holds.} \quad (2.4.1-10)$$

It is clear that the major distinction between the solution to Problem 1 and Problem 3 is a computational one. In Problem 1 the vectors $\{f_1\}_1^n$ provided a ready basis for M . In Problem 3 a basis for S_u must be computed. For example, the set $\{g_1 = (I + F^*F)^{-1}f_1\}_1^n$ is a basis for S_u . Unfortunately the inverse of $(I + F^*F)$ appears not to be solvable in closed form and the element $\{g_1\}_1^n$ must be obtained as solutions to the equations $f_1 = (I + F^*F) g_1$, $1 = 1 \dots n$. A solution technique is discussed in the latter sections (see Sections 3.1 and AX-1).

The element u which minimized $J(u) = \|Fu\|_{H_2}^2 + \|u\|_{H_1}^2$ has been shown to be the unique preimage of x in the set S_u that is

$$(I + F^*F)u = v \text{ for some } v \in N_{F^t} \quad (2.4.1-11)$$

The optimal u can be determined by either of the following two methods:

- (1) Utilizing the steepest descent approximation as indicated in Appendix AX.
- (2) Transforming the integral equation expressed in (2.4.1-11) into a linear system of $2n$ equations. The solution to (2.4.1-11) is stated by the following lemma which will be proved in Section 3.1 of Chapter III.

Lemma (4) Given $x(t_0)$ and $x(t_f)$ in R^n , the element u which is the solution to (2.4.1-11) is given by $u = B^*(t) \lambda(t)$ where $\lambda(t)$ is the solution of a linear system of $2n$ equations

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\lambda}(t) \end{bmatrix} = \begin{bmatrix} A(t) & B(t) B^*(t) \\ I & -A^*(t) \end{bmatrix} \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} \quad (2.4.1-12)$$

with the boundary conditions $x(t_0)=x_0$ and $x(t_f)=x_f$.

2.4.2 Examples

Example 1. Given the dynamic equation

$$\dot{x}(t) = A(t) x(t) + B(t)u(t)$$

find the control function $u(t)$ such that the state of the system transfers from $(0, t_0)$ to (x_f, t_f) while minimizing

$$J(u) = 1/2 \int_{t_0}^{t_f} \{ [x(s), Q(s)x(s)] + [u(s), R(s)u(s)] \} ds$$

where Q and R are positive definite matrices.

The present problem is seemingly more general than the previous case stated in Problem 3. However, if the spaces H_1 and H_2 are defined by

$$H_1 = \left\{ u : \int_{t_0}^{t_f} [u(s), R(s)u(s)] ds < \infty \right\} \quad (2.4.2-1)$$

$$H_2 = \left\{ u : \int_{t_0}^{t_f} [x(s), Q(s)x(s)] ds < \infty \right\} \quad (2.4.2-2)$$

respectively, and the respective inner products on these spaces by the expressions

$$\langle u, v \rangle = \int_{t_0}^{t_f} [u(s), R(s)v(s)] ds \quad u, v \in H_1 \quad (2.4.2-3)$$

$$\langle x, y \rangle = \int_{t_0}^{t_f} [x(s), Q(s)y(s)] ds \quad x, y \in H_2 \quad (2.4.2-4)$$

then $J(u)$ may be written

$$J(u) = 1/2 \|Fu\|_{H_2}^2 + 1/2 \|u\|_{H_1}^2$$

and the present example is included within the framework of Problem 1. It is instructive to note the relatively minor adjustments which must be made in lemma 4 in Section 2.4.1 to treat the present case. Rather than redefining the function spaces, the notation $\langle u, Rv \rangle$ and $\langle x, Qy \rangle$ will be used to denote the right hand sides of the previous equations, thus the complement of the null space of G^t (defined as before), which we shall denote by O , may be computed from

$$\begin{aligned} O &= \left\{ (u, Fu) : \langle u, Rv \rangle_{H_1} + \langle Fu, QFv \rangle_{H_2} = 0, \quad v \in N_{F^t} \right\} \\ &= \left\{ (u, Fu) : \langle (R + F^*QF)u, v \rangle_{H_1} = 0, \quad v \in N_{F^t} \right\} \\ &= \left\{ (u, Fu) : (R + F^*QF)u \in M \right\} \end{aligned} \quad (2.4.2-5)$$

The minimal element may now be determined as before. A basis for S_u , which is defined by

$$S_u = \left\{ u : (R + F^*QF)u \in M \right\}, \quad (2.4.2-6)$$

may be determined by taking the unique solutions $\left\{ g_1 \right\}_1^n$ of

the equations

$$(R + F^*QF)g_i = f_i \quad i = 1, 2, \dots, n \quad (2.4.2-7)$$

where $\{f_i\}_1^n$ is a basis for M . The element u is then the unique linear combination

$$u = \sum_{i=1}^n f_i r_i \quad (2.4.2-8)$$

such that $F^t u = x$ which serves to determine the scalars $r_1 \dots r_n$.

As an alternative approach to the problem we may take advantage of lemma 4 in Section 2.4.1. Using previous arguments the homogeneous differential system defining $\lambda(t)$ and $x(t)$ is given by

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\lambda}(t) \end{bmatrix} = \begin{bmatrix} A(t) & B(t) R^{-1}(t) B^*(t) \\ Q(t) & -A^*(t) \end{bmatrix} \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} \quad (2.4.2-9)$$

The boundary conditions that this system must satisfy are $x(t_0) = x_0$ and $x(t_f) = x_f$.

The solution to equation (2.4.2-9) may be written in partitioned form

$$\begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} = \begin{bmatrix} \mathbb{H}_{11}(t, t_0) & \mathbb{H}_{12}(t, t_0) \\ \mathbb{H}_{21}(t, t_0) & \mathbb{H}_{22}(t, t_0) \end{bmatrix} \begin{bmatrix} x(t_0) \\ \lambda(t_0) \end{bmatrix} \quad (2.4.2-10)$$

It is apparent that the expression $x(t_f) = \mathbb{H}_{11}(t_f, t_0) x(t_0) + \mathbb{H}_{12}(t_f, t_0) \lambda(t_0)$ defines $\lambda(t_0)$. Assuming that $\mathbb{H}_{12}(t_f, t_0)$ is nonsingular we have also from the above,

$$\lambda(t) = \mathbb{H}_{21}(t, t_0) x(t_0) + \mathbb{H}_{22}(t, t_0) \mathbb{H}_{12}^{-1}(t_f, t_0) [x_f - \mathbb{H}_{11}(t_f, t_0) x_0]. \quad (2.4.2-11)$$

Hence the minimizing u^0 is given by

$$u^0(t) = R^{-1}(t)B^*(t) \left\{ \begin{aligned} & \Phi_{21}(t, t_0)x(t_0) \\ & + \Phi_{22}(t, t_0) \Phi_{12}^{-1}(t_f, t_0) [x_f - \Phi_{11}(t_f, t_0)x_0] \end{aligned} \right\}$$

when $x_0=0$ $u^0(t) = R^{-1}(t)B^*(t) \Phi_{22}(t, t_0) \Phi_{12}^{-1}(t_f, t_0)x_f$

(2.4.2-12)

Example 2. Let F^t be a linear bounded transformation of an arbitrary Hilbert function Space H onto R^n find the element $u \in H$ and $x(t) \in R^n$ such that $\psi(x(t_f)) = 0$ and $J(u) = \|x(t_f)\|^2 + \|u\|^2$ is minimized. (2.4.2-13)

This example can be solved by any of the following three methods.

First method: It should be noted that the distinction between the present performance index and the performance index of previous Example 1 is that the notation $\|x(t_f)\|^2$ denotes in the present case a norm on the vector of R^n thus $\|x(t_f)\|^2$ is a function only of the terminal value $(Fu)(t)$.

By replacing F by F^{t_f} and taking the advantage of the solution to Problem 3 of equation (2.4.1-11) the optimal control u for every fixed x is given by

$$u + (F^{t_f})^* F^{t_f} u = B^*(t) \Phi^*(t_0, t) \alpha$$

or
$$u = B^*(t) \Phi^*(t_0, t) \alpha - (F^{t_f})^* x(t_f).$$

It can be shown that

$$(F^{t_f})^* x(t_f) = B^*(t) \Phi^*(t_0, t) \Phi^*(t_f, t_0) x(t_f).$$

Hence

$$u = B^*(t) \Phi^*(t_0, t) [\alpha - \Phi^*(t_f, t_0) x(t_f)].$$

Now since

$$x(t_f) = F^{t_f} u = \Phi(t_f, t_0) \int_{t_0}^{t_f} \Phi^*(t_0, s) B(s) B^*(s) \Phi^*(t_0, s) [\alpha - \Phi^*(t_f, t_0) x(t_f)] ds$$

$$\text{and } x(t_f) = \Phi(t_f, t_0) \beta \quad (2.4.2-14)$$

$$\text{this leads to } \beta = \mathcal{F} [\alpha - \Phi^*(t_f, t_0) x(t_f)],$$

where $\mathcal{F}$ is given by equation (2.2.2-7).

Also the condition $\psi(x(t_f)) = 0$ can be written as

$$\psi(\beta) = 0$$

and the optimal control u is given by

$$u = B^*(t) \Phi^*(t_0, t) \mathcal{F}^{-1} \beta. \quad (2.4.2-15)$$

Substituting equations (2.4.2-14) and (2.4.2-15) in the expression for $J(u)$ in (2.4.2-13).

$$\text{We have } J(\beta) = \langle \beta, \Phi^*(t_f, t_0) \Phi(t_f, t_0) \beta \rangle + \langle \beta, \mathcal{F}^{-1} \beta \rangle \quad (2.4.2-16)$$

The solution to the present example is equivalent to determining the tuple β which minimizes (2.4.2-16) with the constraint $\psi(\beta) = 0$.

Second method: This example can be solved by the application of Problem 2 stated in Section 2.3. Here the function

$$g(x, t_f) = \|x(t_f)\|^2 \quad (2.4.2-17)$$

For every fixed x , the u which minimizes $J(u)$ is given by

$$u = (F^{t_f})^\dagger x(t_f) \quad (2.4.2-18)$$

since $x(t_f)$ has the expansion

$$x(t_f) = \sum_{i=1}^n \phi_i(t_f, t_0) \beta_i \quad (2.4.2-19)$$

Equation (2.4.2-17) reduces to the following form

$$g(x, t_f) = \langle \beta, \Phi^*(t_f, t_0) \Phi(t_f, t_0) \beta \rangle$$

It follows directly from equation (2.3.1-2) that $J(u)$ can be expressed by

$$J(\beta) = \langle \beta, \Phi^*(t_f, t_0) \Phi(t_f, t_0) \beta \rangle + \langle \beta, \mathcal{F}^{-1} \beta \rangle$$

which leads to the same solution given by Method 1.

Third method: Let T_1^t and T_2^t denote the transformations defined respectively by

$$T_1^t : R^n \rightarrow R^n \quad (2.4.2-20)$$

With $T_1^t \triangleq \sum_{i=1}^n \phi_i^+ \langle \phi_i^+ \rangle$ and $\langle \phi_i^+, \phi_j \rangle = \delta_{ij}$ and

$$T_2^t : U \rightarrow R^n \quad (2.4.2-21)$$

with $T_2^t \triangleq \sum_{i=1}^n \phi_i^+ \langle f_i \rangle$

Let the range vector $\left\{ \phi_i^+ \right\}_1^n$ and the functionals on $R^n \left\{ \langle \phi_i^+ \rangle \right\}_1^n$ and on $U \left\{ \langle f_i \rangle \right\}_1^n$ be linearly independent sets; then T_1^t

and T_2^t are linear transformations onto R^n .

Let us introduce a cartesian product space $V = R^n \times U$ and define a transformation T_{12}^t as a direct sum of T_1^t and T_2^t

that is

$$T_{12}^t = T_1^t \oplus T_2^t = \sum_{i=1}^n \phi_i^+ \langle h_i \rangle \quad (2.4.2-22)$$

where $h_i = (\phi_i^+, f_i) \in V$

Since it is clear that T_{12}^t maps V onto R^n then the element

$\omega = (x, u) \in V$ for which $x \in X$ and $u \in U$ has norm $\|\omega\|$

determined by

$$\|\omega\|^2 = \langle \omega, \omega \rangle = \langle (x, u), (x, u) \rangle = \langle x, x \rangle + \langle u, u \rangle = \|x\|^2 + \|u\|^2 \quad (2.4.2-23)$$

The square of norm of the element $\omega \in V$ is, then, precisely the $J(u)$. Thus Example 2 may be reformulated in the following way: consider the linear transformation T_{12}^t mapping V onto R^n with the form

$$T_{12}^t = \sum_{i=1}^n \phi_i \langle h_i$$

where $h_i = (\phi_i^+, f_i) \in V$, the vector $\{\phi_i\}_1^n$ and the functional $\{\langle h_i \rangle\}_1^n$ are linearly independent sets in R^n and V^* respectively.

Find the element ω with minimum norm such that

$$T_{12}^t(\omega) = x(t) \quad (2.4.2-24)$$

By use of the same technique as that given by Problem 1 in Section 2.2.1 the element $\omega \in V$ has the form (see equation 2.2.1-2)

$$\omega = \sum_{j=1}^n h_j \alpha_j \quad (2.4.2-25)$$

Since x is an element in R^n and can be represented by

$$x = \sum_{i=1}^n \phi_i \beta_i \quad (2.4.2-26)$$

the constant α can be determined and expressed as the function of β by the following equation:

$$x = \sum_{i=1}^n \phi_i \beta_i = T_{12}^t(\omega) = \sum_{i=1}^n \phi_i > \langle h_i, \sum_{j=1}^n h_j > \alpha_j \quad (2.4.2-27)$$

This give

$$\beta_i = \sum_{j=1}^n \langle h_i, h_j \rangle \alpha_j \quad (2.4.2-28)$$

because of the independence of $\{\phi_i\}_1^n$.

In vector form we have

$$\begin{aligned} \beta &= H \alpha \\ \text{where} \quad \beta &= \text{col.}(\beta_1 \dots \beta_n) \\ \alpha &= \text{col.}(\alpha_1 \dots \alpha_n) \\ H &= [\langle h_i, h_j \rangle] \end{aligned} \quad (2.4.2-29)$$

It can be shown that

$$H^{-1} = [\langle h_i^+, h_j^+ \rangle]$$

where $\{h_i^+\}_1^n$ denotes the dual basis for V .

Hence we have

$$\alpha = H^{-1} \beta = [\langle h_i^+, h_j^+ \rangle] \beta \quad (2.4.2-30)$$

where in matrix form

$$\begin{aligned} [\langle h_i^+, h_j^+ \rangle] &= [\langle (\phi_i, r_i^+), (\phi_j, r_j^+) \rangle] \\ &= [\langle \phi_i, \phi_j \rangle + \langle r_i^+, r_j^+ \rangle] \\ &= [\Phi^*(t, t_0) \Phi(t, t_0) + \mathcal{F}^{-1}] \end{aligned} \quad (2.4.2-31)$$

where $\mathcal{F} = \int_0^t \Phi(t_0, s) B(s) B^*(s) \Phi^*(t_0, s) ds$ (see equation 2.2.2-7).

Now the minimum energy $\|\omega\|^2$ has the form

$$\begin{aligned} J(u) = \|\omega\|^2 &= \langle \omega, \omega \rangle = \left\langle \sum_{i=1}^n h_i \alpha_i, \sum_{j=1}^n h_j \alpha_j \right\rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \langle h_i, h_j \rangle \alpha_j \\ &= \langle \alpha, \mathcal{H} \alpha \rangle \end{aligned} \quad (2.4.2-32)$$

or in terms of

$$\begin{aligned} J(\beta) &= \langle \beta, \mathcal{F}^{-1} \beta \rangle \\ &= \langle \beta, [\Phi^*(t, t_0) \Phi(t, t_0) + \mathcal{F}^{-1}] \beta \rangle \\ &= \langle \beta, \Phi^*(t, t_0) \Phi(t, t_0) \beta \rangle + \langle \beta, \mathcal{F}^{-1} \beta \rangle \end{aligned} \quad (2.4.2-33)$$

The desired $x \in R^n$ and $u \in U$ can now be computed by the equation (2.4.2-25), that is

$$\omega = \sum_{i=1}^n h_i \alpha_i$$

or

$$\omega = \sum_{i=1}^n h_i^+ \beta_i = \sum_{i=1}^n (\phi_i, r_i^+) \beta_i$$

which gives

$$x = \sum_{i=1}^n \phi_i \beta_i \quad (2.4.2-34)$$

and

$$u = \sum_{i=1}^n r_i^+ \beta_i \quad (2.4.2-35)$$

The solution to this example reduces to the determination of β which minimizes (2.4.2-33) with the constraint $\psi(\beta) = 0$.

Example 3. Application of Problem 3 to the open loop system reliability can be formulated in the following.

Given the continuous linear dynamic system governed by

$$\dot{x}(t) = A(t, a, \xi) x(t) + B(t) u(t), \quad t \in \mathcal{T} = [t_0, t_f] \quad (2.4.2-36)$$

and $x(t_0) = x_0$, $E[x(t_f)] = \bar{x}_f$ where (x_0, t_0) is a fixed point, a , $c_1 \leq a \leq c_2$, is the random variable with known statistical properties, c_1 and c_2 are two constants, and ξ is the unknown parameter to be determined.

Find the control function $u(t)$ and the parameter ξ such that the index of performance

$$J = \frac{1}{2} E \|\delta x\|^2 + \frac{1}{2} \int_{t_0}^{t_f} [u(s), R(s)u(s)] ds \quad (2.4.2-37)$$

is minimized where $E[y]$ denotes the expectation of y .

Assumptions:

(1) The matrix $A(t, a, \xi)$ which is subjected to a random variation is expressed by

$$A = \bar{A} + \delta A \quad (2.4.2-38)$$

where $\bar{A}$ = mean value of A

δA = deviation of A from its mean value.

(2) The variation in δA is small with respect to A (higher order terms in δA are negligible) so that the stability of the system will not be affected by the change in δA .

(3) The vector $\xi = (\xi_1 \dots \xi_n)$ is allowed to range over the set of values defined by

$$\Gamma = \left\{ \xi : \xi_1^2 + \xi_2^2 + \dots + \xi_n^2 \leq c^2 \right\} \text{ for some fixed constant } c.$$

(4) R is symmetric, positive definite, and is generally a time varying $m \times m$ matrix.

(5) The matrix D defined by

$$D = \begin{bmatrix} \frac{\partial^2 J^0}{\partial \xi_1 \partial \xi_1} & \dots & \frac{\partial^2 J^0}{\partial \xi_1 \partial \xi_n} \\ \vdots & & \\ \frac{\partial^2 J^0}{\partial \xi_n \partial \xi_1} & \dots & \frac{\partial^2 J^0}{\partial \xi_n \partial \xi_n} \end{bmatrix} \quad (2.4.2-39)$$

is positive definite, where J^0 is optimal index of performance with respect to the optimal control function u^0 , i.e., it is the value evaluated at u^0 .

Solution: If the random variation in A ($A = \bar{A} + \delta A$) causes the deviation of the trajectory from its nominal path, $\bar{x}$, by δx where the state of the system $x = \bar{x} + \delta x$, then the equation of motion of the nominal trajectory is given by

$$\dot{\bar{x}} = \bar{A}\bar{x} + Bu \quad (2.4.2-40)$$

and the perturbation of the trajectory, neglecting second order terms in δA and δx , is described by

$$\dot{\delta x} = \bar{A} \delta x + \delta A \bar{x} \quad (2.4.2-41)$$

It is well known that the solution to (2.4.2-40) and (2.4.2-41) are respectively expressible in the following matrix integral form

$$\bar{x}(t, t_0) = \Phi(t, t_0)x_0 + \int_{t_0}^t \Phi(t, s)B(s)u(s)ds \quad (2.4.2-42)$$

$$\delta x(t, t_0) = \Phi(t, t_0)\delta x_0 + \int_{t_0}^t \Phi(t, \tau)\delta A(\tau)\bar{x}(\tau) d\tau \quad (2.4.2-43)$$

where δx and δA are function of random variable a . Since (x_0, t_0) is fixed without error that is $\delta x_0 = 0$, at $t = t_0$ equation (2.4.2-43) reduces to

$$\delta x(t, t_0) = \int_{t_0}^t \Phi(t, \tau) \delta A(\tau) \bar{x}(\tau) d\tau \quad (2.4.2-44)$$

In many practical problems the functional J may be extremely complex and thus causing difficulties in analyzing its extremization. Assume however that a second functional I may be found which is simple in form and bounds J in the sense that

$$J \leq MI \quad (2.4.2-45)$$

for all parameter values.

In this situation we are tempted to replace the problem of finding minimum J by the problem of seeking minimum I . Since these are two different problems, we obtain only an approximate solution to the first. The approximate solution may be used as an estimate of the true optimum or perhaps as a first estimate in the optimization procedure. Such an approach reduces significantly in the difficulties in calculation. This technique is frequently used in practical problems such as the one given in reference (8). By means of the Schwarz inequality it can be shown in the Appendix AIX that

$$\|\delta x\|^2 \leq M^2 \int_{t_0}^t [\bar{x}^* \delta A^* \delta A \bar{x}] ds ; \quad (2.4.2-46)$$

hence

$$E \|\delta x\|^2 \leq \int_{t_0}^t [\bar{x}, Q \bar{x}] ds$$

where V^* denotes the transpose of V

$$Q = M^2 E [\delta A^* \delta A], \quad (2.4.2-47)$$

and

$$M^2 = \int_{t_0}^{t_f} \int_{t_0}^t \sum_{i,k=1}^n W_{ik}^2(t, \tau) d\tau dt \quad (2.4.2-48)$$

This M so obtained is the smallest number for which the above inequalities hold, and W_{ik} = the element of the matrix $\Phi(t, \tau)$. Now if A is a constant matrix, then $\Phi(t, \tau) = \Phi(t - \tau)$. We shall solve the above problem by the following approach.

(1) Fix ξ and find the optimal $u(t)$ as a function of ξ such that

$$\begin{aligned} J &= \frac{1}{2} E \|\delta x\|^2 + \frac{1}{2} \int_{t_0}^{t_f} [u, Ru] ds \\ &\triangleq \frac{1}{2} \int_{t_0}^{t_f} \left\{ [\bar{x}, Q \bar{x}] + [u, Ru] \right\} ds \end{aligned} \quad (2.4.2-49)$$

has a minimum value (denoted by $J^0(\xi)$).

(2) Choose ξ such that $J^0(\xi)$ attains a minimum value subject to the following inequality constraint:

$$\xi_1^2 + \xi_2^2 + \dots + \xi_n^2 \leq c^2$$

where c is a constant. Note that once $u(t)$ is found by substituting in (2.4.2-49) it is clear that $J^0(\xi)$ is a function of $\xi = (\xi_1 \dots \xi_n)$ only.

To formulate this problem in the notation of main discussion in Section 2.4.1, we introduce the space H_1 , H_2 and Γ defined by

$$H_1 = \left\{ u : \int_{t_0}^t [u(s), R(s)u(s)] ds < \infty \right\},$$

$$H_2 = \left\{ x : \int_{t_0}^t [x(s), Q(s)u(s)] ds < \infty \right\},$$

$$\Gamma = \left\{ \xi : \xi_1^2 + \dots + \xi_n^2 \leq c^2 \right\}$$

respectively, and the respective inner product on H_1 and H_2 spaces by expressions

$$\langle u, v \rangle = \int_{t_0}^{t_f} [u(s), R(s)u(s)] ds \quad u, v \in H$$

$$\langle x, y \rangle = \int_{t_0}^{t_f} [x(s), Q(s)y(s)] ds \quad x, y \in H$$

By the application of Example 1 of Problem 3 in Section 2.4.2 to the above example, the optimal $u(t)$ for fixed ξ can be found as given by Equation (2.4.2-12). The optimal index $J^0(\xi)$ can be obtained by equation (3.5-7). Finally the optimal parameter ξ can be determined by minimizing $J^0(\xi)$ subject to the constraint $\xi_1^2 + \dots + \xi_n^2 \leq c^2$.

It should be noted that the variation of $\xi = (\xi_1 \dots \xi_n)$

in this example is small over the set defined by

$$\Gamma = \left\{ \xi : \xi_1^2 + \dots + \xi_n^2 \leq c^2 \right\} \text{ such that } J^0(\xi) \text{ has a convex}$$

property within the set Γ . Therefore the method of steepest descent is applicable to the above example for the

determination of ξ_1 ($i=1,2,\dots,n$).

Let us give a numerical computation to illustrate a class of physical problems giving rise to the problem of system reliability. Given a dynamic plant described by the following block diagram

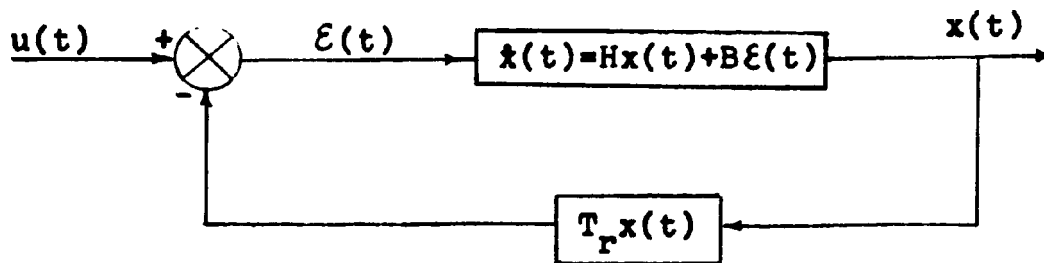


Fig. 2.4.2-1.--A feedback control system with random variable "a" and unknown parameter $\xi = (\xi_1, \xi_2)$.

assume that

$$H = \begin{bmatrix} -a & 0 \\ 0 & -a \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$T_r = \begin{bmatrix} \xi_1 & 0 \\ 0 & \xi_2 \end{bmatrix}.$$

(2.4.2-50)

The random variable "a" has the probability density function given by

$$f(a) = 2 \quad 0 \leq a \leq \frac{1}{2}$$

$$= 0 \quad \text{otherwise.}$$

The problem is to find the parameter (ξ_1, ξ_2) with the

constraint

$$\xi_1^2 + \xi_2^2 \leq \left(\frac{1}{8}\right)^2$$

and the control function $u(t) = (u_1(t), u_2(t))$ for given $x(0)=x_0$ and $E[x(t_1)]=0$ such that the system attains a highest possible reliability with minimum input energy.

We may transform this problem into an equivalent form stated in the following.

To choose $u(t) = (u_1(t), u_2(t))$, $0 \leq t \leq t_1$

and $\xi = (\xi_1, \xi_2)$

such that

$$J = \frac{1}{2} E \| \delta x \|^2 + \frac{1}{2} \int_0^{t_1} [u(s), R(s)u(s)] ds$$

takes on the least possible value with u, ξ, x subject to the following constraints.

$$(1) \quad \dot{x}(t) = A x(t) + B u(t), \quad 0 \leq t \leq t_1$$

$$(2) \quad \xi_1^2 + \xi_2^2 \leq \left(\frac{1}{8}\right)^2$$

$$(3) \quad x(0) = x_0, \quad \bar{x}(t_1) = 0$$

$$(4) \quad A = H - T_r = \bar{A} + \delta A$$

where

$$\bar{A} = - \begin{bmatrix} \bar{a} + \xi_1 & 0 \\ 0 & \bar{a} + \xi_2 \end{bmatrix} \quad \text{and} \quad \delta A = - \begin{bmatrix} \delta a & 0 \\ 0 & \delta a \end{bmatrix}$$

In the above equation $a = \bar{a} + \delta a$, $\bar{a}$ is the mean value of the random variable "a." The probability density of a is given

$$\begin{aligned} \text{by} \quad f(a) &= 2 & 0 \leq a \leq \frac{1}{2} \\ &= 0 & \text{otherwise,} \end{aligned}$$

and $B = R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

First we compute $E[\delta A^* \delta A]$, M^2 and Q in terms of ξ_1 and ξ_2 (or r_1 and r_2 where $r_1 = \bar{a} + \xi_1$ and $r_2 = \bar{a} + \xi_2$).

(1) Computation of $E[\delta A^* \delta A]$:

first find $\bar{a} = E[a] = \int_0^{1/2} af(a)da = 1/4$ (2.4.2-51)

now $f(\delta a) = f(a - \bar{a}) = f(a - 1/4) = 2, \quad -1/4 \leq a \leq 1/4$

hence $E[\delta a^2] = \int_{-1/4}^{1/4} \lambda^2 f(\lambda) d\lambda = 1/48, \quad (2.4.2-52)$

then

$$E[\delta A^* \delta A] = E\left\{ \begin{bmatrix} -\delta a & 0 \\ 0 & -\delta a \end{bmatrix} \begin{bmatrix} -\delta a & 0 \\ 0 & -\delta a \end{bmatrix} \right\} = \begin{bmatrix} 1/48 & 0 \\ 0 & 1/48 \end{bmatrix}$$

Also since

$$\bar{A} = - \begin{bmatrix} \xi_1 + 1/4 & 0 \\ 0 & \xi_2 + 1/4 \end{bmatrix}$$

by change of variable $\xi_1 + 1/4 = r_1 \quad (i=1, 2)$

then

$$\bar{A} = - \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}$$

and the constraint equation becomes

$$(r_1 - 1/4)^2 + (r_2 - 1/4)^2 \leq (1/8)^2 \quad (2.4.2-53)$$

(2) Computation of M^2

Since $W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} = \begin{bmatrix} e^{-r_1(t-\tau)} & 0 \\ 0 & e^{-r_2(t-\tau)} \end{bmatrix},$

we have

$$\begin{aligned} M^2 &= \int_0^{t_1} \int_0^t (w_{11}^2 + w_{12}^2 + w_{21}^2 + w_{22}^2) d\tau dt \\ &= \int_0^{t_1} \int_0^t [e^{-2r_1(t-\tau)} + e^{-2r_2(t-\tau)}] d\tau dt \end{aligned}$$

thus

$$M^2(r_1, r_2) = \frac{e^{-2r_1 t_1} - 1 + 2 r_1 t_1}{4r_1^2} + \frac{e^{-2r_2 t_1} - 1 + 2 r_2 t_1}{4r_2^2} \quad (2.4.2-54)$$

(3) Computation of Q

$$\text{Now } Q = M^2 E [\delta \bar{A}^* \delta A] = \begin{bmatrix} \frac{M^2}{48} & 0 \\ 0 & \frac{M^2}{48} \end{bmatrix} = \begin{bmatrix} M_1 & 0 \\ 0 & M_1 \end{bmatrix} \quad (2.4.2-55)$$

$$\text{where } M_1 = \frac{M^2}{48} \quad (2.4.2-56)$$

and M is given by (2.4.2-54).

To find the optimal $u(t)$ we shall compute

$$\hat{H} = \begin{bmatrix} \hat{H}_{11}(t, t_0) & \hat{H}_{12}(t, t_0) \\ \hat{H}_{21}(t, t_0) & \hat{H}_{22}(t, t_0) \end{bmatrix}$$

that satisfies

$$\dot{\hat{H}}(t, t_0) = Z \hat{H}(t, t_0)$$

and

$$\hat{H}_{11}(t_0, t_0) = \hat{H}_{22}(t_0, t_0) = I, \quad \hat{H}_{12}(t_0, t_0) = \hat{H}_{21}(t_0, t_0) = 0$$

where

$$Z = \begin{bmatrix} \bar{A} & B R B^* \\ Q & -\bar{A}^* \end{bmatrix} \quad (2.4.2-57)$$

$\bar{A}$ denotes the mean value of A.

The result of the submatrices $\textcircled{H}_{11}$, $\textcircled{H}_{12}$, $\textcircled{H}_{21}$, and $\textcircled{H}_{22}$ are easily obtained and expressible in the following form:

$$\textcircled{H}_{11}(t, t_0) = \begin{bmatrix} \cosh \sqrt{r_1^2 + M_1} t - \frac{r_1}{\sqrt{r_1^2 + M_1}} \sinh \sqrt{r_1^2 + M_1} t & 0 \\ 0 & \cosh \sqrt{r_2^2 + M_1} t - \frac{r_2}{\sqrt{r_2^2 + M_1}} \sinh \sqrt{r_2^2 + M_1} t \end{bmatrix}$$

$$\textcircled{H}_{12}(t, t_0) = \begin{bmatrix} \frac{\sinh \sqrt{r_1^2 + M_1} t}{\sqrt{r_1^2 + M_1}} & 0 \\ 0 & \frac{\sinh \sqrt{r_2^2 + M_1} t}{\sqrt{r_2^2 + M_1}} \end{bmatrix}$$

$$\textcircled{H}_{21}(t, t_0) = \begin{bmatrix} \frac{-r_1}{\sqrt{r_1^2 + M_1}} \sinh \sqrt{r_1^2 + M_1} t & 0 \\ 0 & \frac{-r_2}{\sqrt{r_2^2 + M_1}} \sinh \sqrt{r_2^2 + M_1} t \end{bmatrix}$$

and

$$\textcircled{H}_{22}(t, t_0) = \begin{bmatrix} \cosh \sqrt{r_1^2 + M_1} t + \frac{r_1}{\sqrt{r_1^2 + M_1}} \sinh \sqrt{r_1^2 + M_1} t & 0 \\ 0 & \cosh \sqrt{r_2^2 + M_1} t + \frac{r_2}{\sqrt{r_2^2 + M_1}} \sinh \sqrt{r_2^2 + M_1} t \end{bmatrix}$$

where $t_0 = 0$ in this example.

The optimal control function $u^*(t)$ with fixed ξ is given by:

$$u^*(t) = R^{-1} B^* [\bar{H}_{21}(t, t_0) - \bar{H}_{22}(t, t_0) \bar{H}_{12}^{-1}(t_1, t_0) \bar{H}_{11}(t_1, t_0)] x(t_0) \quad (2.4.2-58)$$

The control function $u^*(t)$ can be expressed as a function of the instantaneous value of x by replacing t_0 by t in (2.4.2-58). This yields

$$u^*(t) = -R^{-1} B^* \bar{H}_{12}^{-1}(t_1, t) \bar{H}_{11}(t_1, t) x(t) = -k(t) x(t), \quad (2.4.2-59)$$

$$\text{where } k(t) = R^{-1} B^* \bar{H}_{12}^{-1}(t_1, t) \bar{H}_{11}(t_1, t)$$

is the optimal gain equation. Equation (2.4.2-59) represents the control law for the system shown in Figure 2.4.2-1. The optimal index of performance can be obtained from equation (3.5-7)

$$\begin{aligned} J^* &= -1/2 \langle x(t_0), \lambda(t_0) \rangle \\ &= 1/2 \langle x(t_0), \bar{H}_{12}^{-1}(t_1, t_0) \bar{H}_{11}(t_1, t_0) x(t_0) \rangle \end{aligned}$$

which gives

$$J^* = \frac{x_{10}^2}{2} \left[\sqrt{r_1^2 + M_1} \coth \sqrt{r_1^2 + M_1} t_1 - r_1 \right] + \frac{x_{20}^2}{2} \left[\sqrt{r_2^2 + M_1} \coth \sqrt{r_2^2 + M_1} t_1 - r_2 \right] \quad (2.4.2-60)$$

where M_1 is given by (2.4.2-56)

$$\text{and } (r_1 - 1/4)^2 + (r_2 - 1/4)^2 \leq (1/8)^2$$

Fortunately the rather cumbersome expression of (2.4.2-60) has a simple approximate form in the set

$$\Gamma = \left\{ r: (r_1 - 1/4)^2 + (r_2 - 1/4)^2 \leq (1/8)^2 \right\}.$$

For convenience, we shall assume that $x_{10} = x_{20} = \sqrt{2}$ and $t_1 = 1$. Expanding the function e^y and $y \coth y$ using fourth order terms in y will give a good approximation for small or modest sizes of y . By doing so we have

$$M_1 = 1/48 \left(1 - \frac{r_1 + r_2}{3} + \frac{r_1^2 + r_2^2}{6} \right)$$

$$\begin{aligned} \text{and } \sqrt{r_1^2 + M_1} \coth \sqrt{r_1^2 + M_1} = & 1 + 1/3[r_1^2 + 1/48(1 - \frac{r_1 + r_2}{3} + \frac{r_1^2 + r_2^2}{6})] \\ & - 1/45[r_1^2 + 1/48(1 - \frac{r_1 + r_2}{3} + \frac{r_1^2 + r_2^2}{6})] \end{aligned} \quad (2.4.2-61)$$

since the parameters r_1 is within the limit $1/8 \leq r_1 \leq 3/8$ (see equation 2.4.2-53). The upper bounded value of $\sqrt{r_1^2 + M_1} \coth \sqrt{r_1^2 + M_1}$ for $r_1 = r_2 = 3/8$ in equation (2.4.2-61) is given by $1 + 0.0525 - 0.000552$. Clearly the third term is much smaller than the first and second term in equation (2.4.2-61) for the parameter r_1 ranged over the interval $1/8 \leq r_1 \leq 3/8$. The analysis is considerably simplified by neglecting this third term. Then total expression of J° is given by

$$\begin{aligned} J^\circ = & 2 + 1/3[r_1^2 + 1/48(1 - \frac{r_1 + r_2}{3} + \frac{r_1^2 + r_2^2}{6})] - r_1 \\ & + 1/3[r_2^2 + 1/48(1 - \frac{r_1 + r_2}{3} + \frac{r_1^2 + r_2^2}{6})] - r_2 \\ \text{or } J^\circ = & 0.3356r_1^2 + 0.3356r_2^2 - 1.0046r_1 - 1.0046r_2 + 2.014 \end{aligned} \quad (2.4.2-62)$$

From the numerical comparisons of J° obtained from (2.4.2-60) and (2.4.2-62) which are listed in Table 2.4.2-1, it can be seen that equation (2.4.2-62) does represent an

excellent and simple approximation to equation (2.4.2-60)
for $1/8 \leq r_1 \leq 3/8$ ($i=1,2$), $x_{10}=x_{20}=\sqrt{2}$ and $t_1=1$.

TABLE 1

SOME NUMERICAL COMPARISONS OF J°
OBTAINED FROM EQUATIONS
(2.4.2-60) AND (2.4.2-62)

Equation	$r_1=r_2=3/8$	$r_1=r_2=2/8$	$r_1=r_2=1/8$	$r_1=1/8, r_2=2/8$
J° app. Calculated from 2.4.2-62	1.3552	1.5536	1.7733	1.564
J° true Calculated from 2.4.2-60	1.3522	1.5508	1.7712	1.561
Error	0.22%	0.18%	0.12%	0.19%

From equation (2.4.2-62) it can be readily shown that

$$\frac{\partial^2 J^\circ}{\partial r_1^2} \geq 0$$

and

$$\det. \begin{bmatrix} \frac{\partial^2 J^\circ}{\partial r_1^2} & \frac{\partial^2 J^\circ}{\partial r_1 \partial r_2} \\ \frac{\partial^2 J^\circ}{\partial r_2 \partial r_1} & \frac{\partial^2 J^\circ}{\partial r_2^2} \end{bmatrix} \geq 0$$

for all $1/8 \leq r_1 \leq 3/8$, $i=1,2$

Hence the Hessian matrix, which is defined by

$$\left[\frac{\partial^2 J^0}{\partial r_1 \partial r_j} \right] \quad 1, j = 1, 2$$

is a positive definite matrix. This leads to the fact that $J^0(r_1, r_2)$ is a convex function of r_1 and r_2 within the set Γ , defined by

$$\Gamma = \left\{ r : (r_1 - 1/4)^2 + (r_2 - 1/4)^2 \leq (1/8)^2 \right\}$$

Since $J^0(r_1, r_2)$ in (2.4.2-62) is a twice continuous differentiable function with respect to r_1 and r_2 and $g(r) = (r_1 - 1/4)^2 + (r_2 - 1/4)^2 - (1/8)^2$ is a convex function, the method of steepest descent can be applied to the above problem in order to determine the parameters r_1 and r_2 in the minimization of J^0 subjected to the inequality constraint $g(r) \leq 0$.

2.5 Problem 4

Let H_1 and H_2 be two Hilbert function spaces and F be a transfer operator carrying H_1 into H_2 . Let F^t be the sample of F . If $\hat{u}$ and $\hat{y}$ denote arbitrary elements of H_1 and H_2 respectively from the set $U_x = \{u : u \in H_1, F^t u = x\}$ find the element which minimizes the functional

$$J(u) = \left\| Fu - \hat{y} \right\|_{H_2}^2 + \left\| u - \hat{u} \right\|_{H_1}^2.$$

In Problem 4 the present line of development is brought to its full potential as far as theoretical considerations are concerned. We shall use as before the notation $H(F)$ for the graph of F . The problem may be split into two cases.

2.5.1 Abstract Theory

Case I: In general the pair $(\hat{u}, \hat{y})$ is an arbitrary element of $H_1 \times H_2$. For the present we restrict attention to the case where $(\hat{u}, \hat{y}) \in H(F)$. Under these conditions theorem C applies (replacing $(F^t)^\dagger$ by $(G^t)^\dagger$). Indeed let (u°, Fu°) be defined by

$$(u^\circ, Fu^\circ) = (G^t)^\dagger [x - \hat{x}] + (\hat{u}, F\hat{u}) \quad (2.5.1-1)$$

where $\hat{x}$ is defined by $\hat{x} = F^t \hat{u} = \hat{y}(t)$ and the property $\hat{y} = F\hat{u}$ has been employed. Operating with G^t on both sides of this equation yields

$$\begin{aligned} G^t(u^\circ, Fu^\circ) &= F^t u^\circ = x - \hat{x} + G^t(\hat{u}, F\hat{u}) \\ &= x - \hat{x} + F^t \hat{u} = x \end{aligned}$$

hence (u°, Fu°) is a preimage of x . Conversely the equality

$$(u^\circ, Fu^\circ) - (\hat{u}, F\hat{u}) = (u^\circ - \hat{u}, Fu^\circ - F\hat{u}) = (G^t)^\dagger [x - \hat{x}]$$

implies by the definition of $(G^t)^\dagger$ that $(u^\circ - \hat{u}, Fu^\circ - F\hat{u})$ is the element in $H(F)$ with minimum norm mapping into $x - \hat{x}$.

However since

$$\|(u^\circ - \hat{u}, Fu^\circ - F\hat{u})\|_{H(F)}^2 = \|u^\circ - \hat{u}\|_{H_2}^2 + \|Fu^\circ - F\hat{u}\|_{H_1}^2 = J(u^\circ)$$

and $\hat{u}, \hat{y}, \hat{x}$, are all fixed it is clear that u° is the solution to Problem 4 for Case I.

Equation (2.5.1-1) may be rewritten as

$$G^t(u^\circ - \hat{u}, F(u^\circ - \hat{u})) = x - \hat{x} \quad (2.5.1-2)$$

which has the same form as equation (2.4.1-4). By the application of theorem E, the optimal u° is satisfied by

$$(I + F^*F)(u^\circ - \hat{u}) \in v \quad \text{for } v \in N_{F^t} \quad (2.5.1-3)$$

The computational technique for finding u^0 is essentially the same as that of Section 3.1 (or lemma 4 in Section 2.4.1).

Case II: Let $(\hat{u}, \hat{y})$ be an arbitrary element of $H_1 \times H_2$.

Since $H(F)$ is a closed linear subspace of $H_1 \times H_2$ it follows that the decomposition

$$H_1 \times H_2 = H(F) \oplus H(F)^\perp \quad (2.5.1-4)$$

may be defined along with the orthogonal projection P_H of $H_1 \times H_2$ on $H(F)$. Let $(\bar{u}, \bar{y})$ be defined by

$$(\bar{u}, \bar{y}) \triangleq P_H(\hat{u}, \hat{y}) \quad (2.5.1-5)$$

then $(\bar{u}, \bar{y})$ is in $H(F)$ and $(\hat{u}, \hat{y}) - (\bar{u}, \bar{y}) \in H(F)^\perp$ and thus the equalities

$$\begin{aligned} \|Fu - \hat{y}\|_{H_2}^2 + \|u - \hat{u}\|_{H_1}^2 &= \|(Fu - \bar{y}) + (\bar{y} - \hat{y})\|_{H_2}^2 + \|(u - \bar{u}) + (\bar{u} - \hat{u})\|_{H_1}^2 \\ &= \|Fu - \bar{y}\|_{H_2}^2 + \|\bar{y} - \hat{y}\|_{H_2}^2 + \|u - \bar{u}\|_{H_1}^2 + \|\bar{u} - \hat{u}\|_{H_1}^2 \\ &= \|Fu - \bar{y}\|_{H_2}^2 + \|u - \bar{u}\|_{H_1}^2 + \|\bar{y} - \hat{y}\|_{H_2}^2 + \|\bar{u} - \hat{u}\|_{H_1}^2 \end{aligned} \quad (2.5.1-6)$$

hold for all $u \in H_1$. The last two terms are independent of u and hence it is apparent that Case II is reduced to Case I by taking the projection of $(\hat{u}, \hat{y})$ on the graph of F .

The projection P_H : Let the operator V on the Hilbert space $H_1 \times H_2$ be defined by the relation

$$V(u, v) = (-v, u), \quad (u, v) \in H_1 \times H_2 \quad (2.5.1-7)$$

It can be proved (1) (also see Appendix AXI) that the

orthogonal complement of $H(F)$ in $H_1 \times H_2$ is given by

$$H(F)^\perp = VH(F^*) \quad (2.5.1-8)$$

A consequence of this result is that the orthogonal decomposition of $H_1 \times H_2$ on $H(F)$ is given explicitly by the formula

$$H_1 \times H_2 = H(F) \oplus VH(F^*) \quad (2.5.1-9)$$

This formula will prove useful in defining the orthogonal projection of $H_1 \times H_2$ on $H(F)$.

Let $(\hat{u}, \hat{y}) \in H_1 \times H_2$ denote an arbitrary vector. Then in view of equation (2.5.1-9) a unique decomposition exists of the form

$$\begin{aligned} (\hat{u}, \hat{y}) &= (\bar{u}, F\bar{u}) + V(u_1, F^*u_1) \\ &= (\bar{u}, F\bar{u}) + (-F^*u_1, u_1) \\ &= (\bar{u} - F^*u_1, F\bar{u} + u_1) \end{aligned} \quad (2.5.1-10)$$

This equality implies the two equalities

$$\begin{aligned} \hat{u} &= \bar{u} - F^*u_1 \\ \hat{y} &= F\bar{u} + u_1 \end{aligned} \quad (2.5.1-11)$$

Operating on the second equation with F^* and using the result to eliminate F^*u_1 in the first equation produces the result

$$\bar{u} = (I + F^*F)^{-1} [\hat{u} + F^*\hat{y}] \quad (2.5.1-12)$$

Similarly it is easily shown that u_1 satisfies

$$u_1 = (I + F^*F)^{-1} [\hat{y} - F\hat{u}] \quad (2.5.1-13)$$

Since the pair $(\bar{u}, F\bar{u})$ is determined completely by $\bar{u}$ equations, (2.5.1-12) is sufficient to define the projection P_H of $H_1 \times H_2$ on $H(F)$.

2.5.2 Examples

Example 1: Given a plant whose dynamic characteristic is defined by

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \text{ where } t \in \mathcal{T} = [t_0, t_f] \quad (2.5.2-1)$$

and the reference trajectory $\hat{x}(t)$.

Find the control function u such that $x(t_0)$ transfers from $x(t_0) = 0$ to $x(t_f) = x_f$ and

$$J(u) = 1/2 \int_{t_0}^{t_f} \left\{ [(x - \hat{x}), (x - \hat{x})] + [u, u] \right\} ds \quad (2.5.2-2)$$

is minimized.

Solution: We may reformulate the above problem in the notation used in Problem 4, namely, given $x \in R^n$ and let F^t be the transformation from U onto R^n , to find the element $u \in U$ such that

$$J(u) = 1/2 \|x - \hat{x}\|_{H_2}^2 + 1/2 \|u\|_{H_1}^2 \quad (2.5.2-3)$$

is minimized.

For the present case $\hat{u} = 0$ and $(\hat{x}, 0)$ is a tuplet of $H_1 \times H_2$. The relation $F\hat{x} = \hat{u}$ does not hold. Case 2, therefore, in problem 4 is considered. Replacing $\hat{u}$ in equation (2.5.1-3) by $\bar{u}$ expressed in equation (2.5.1-2) gives:

$$(I + F^*F) [u^0 - (I + F^*F)^{-1}(\hat{u} + F^*\hat{x})] \in v \quad \text{for } v \in N_{F^t} \quad (2.5.2-4)$$

$$\text{or } (I + F^*F)u^0 - (\hat{u} + F^*\hat{x}) = B^*(t) \Phi^*(t_0, t) c \quad (2.5.2-5)$$

$$\text{or } u^0 - \hat{u} = B^*(t) \Phi^*(t_0, t) \left\{ c - \int_{t_0}^{t_f} \Phi^*(s, t_0) [x(s) - \hat{x}(s)] ds \right\} \quad (2.5.2-6)$$

Utilizing the same argument as stated in the lemma 4 given by Section 2.4.1, it can be shown that the optimal u^0 is given by

$$u^0 - \hat{u} = B^*(t) \lambda(t) \quad (2.5.2-7)$$

where $\lambda(t)$ is the solution to the following $2n$ linear differential equation:

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\lambda}(t) \end{bmatrix} = \begin{bmatrix} A(t) & B(t)B^*(t) \\ I & -A^*(t) \end{bmatrix} \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} + \begin{bmatrix} B(t)\hat{u}(t) \\ -\hat{x}(t) \end{bmatrix} \quad (2.5.2-8)$$

The solution can be written as

$$\begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} = \begin{bmatrix} \mathcal{H}_{11}(t, t_0) & \mathcal{H}_{12}(t, t_0) \\ \mathcal{H}_{21}(t, t_0) & \mathcal{H}_{22}(t, t_0) \end{bmatrix} \begin{bmatrix} x(t_0) \\ \lambda(t_0) \end{bmatrix} + \begin{bmatrix} \mathcal{J}_1(t, t_0) \\ \mathcal{J}_2(t, t_0) \end{bmatrix} \quad (2.5.2-9)$$

where

$$\begin{bmatrix} \mathcal{J}_1(t, t_0) \\ \mathcal{J}_2(t, t_0) \end{bmatrix} = \int_{t_0}^t \begin{bmatrix} \mathcal{H}_{11}(t, s) & \mathcal{H}_{12}(t, s) \\ \mathcal{H}_{21}(t, s) & \mathcal{H}_{22}(t, s) \end{bmatrix} \begin{bmatrix} B(s)\hat{u}(s) \\ -\hat{x}(s) \end{bmatrix} ds \quad (2.5.2-10)$$

Now when $\hat{u} = 0$

$$\begin{aligned} \mathcal{J}_1(t, t_0) &= - \int_{t_0}^t \mathcal{H}_{12}(t, s) \hat{x}(s) ds \\ \mathcal{J}_2(t, t_0) &= - \int_{t_0}^t \mathcal{H}_{22}(t, s) \hat{x}(s) ds \end{aligned} \quad (2.5.2-11)$$

From the boundary conditions $x(t_0) = x_0$ and $x(t_f) = x_f$ we have

$$\lambda(t_0) = \mathcal{H}_{12}^{-1}(t_f, t_0) [x(t_f) - \mathcal{H}_{11}(t_f, t_0)x_0 - \mathcal{J}_1(t_f, t_0)] \quad (2.5.2-12)$$

Substituting (2.5.1-12) in the second equation of (2.5.1-9) this gives

$$\lambda(t) = \mathbb{H}_{21}(t, t_0)x_0 + \mathbb{H}_{22}(t, t_0) \mathbb{H}_{12}^{-1}(t_f, t_0)[x_f - \mathbb{H}_{11}(t_f, t_0)x_0 - \int_1(t_f, t_0)] + \int_2(t, t_0)$$

when $x(t_0) = x_0 = 0$

$$\lambda(t) = \mathbb{H}_{22}(t, t_0) \mathbb{H}_{12}(t_f, t_0)^{-1}[x_f - \int_1(t_f, t_0)] + \int_2(t, t_0)$$

(2.5.2-13)

Substituting (2.5.2-13) in (2.5.2-7) and setting $\hat{u} = 0$ the optimal u^0 is given by

$$u^0(t) = B^*(t) \left\{ \mathbb{H}_{22}(t, t_0) \mathbb{H}_{12}(t_f, t_0)^{-1}[x_f - \int_1(t_f, t_0)] + \int_2(t, t_0) \right\}$$

If the final value is $x(t_f) = \hat{x}(t_f)$, then the optimal control function is

$$u^0(t) = B^*(t) \left\{ \mathbb{H}_{22}(t, t_0) \mathbb{H}_{12}(t_f, t_0)^{-1}[\hat{x}(t_f) - \int_1(t_f, t_0)] + \int_2(t, t_0) \right\}$$

(2.5.2-14)

Remarks: If the index of performance takes the following

form

$$J(u) = 1/2 \int_{t_0}^{t_f} \left\{ [(x - \hat{x}), Q(x - \hat{x})] + [u, Ru] \right\} ds \quad (2.5.2-15)$$

where Q and R are symmetric positive definite matrices, then following the same argument as stated in Example 1 of Section 2.4.2 by introducing

$$\langle u, v \rangle = \int_{t_0}^{t_f} [u(s), R(s) u(s)] ds$$

$$\langle x, y \rangle = \int_{t_0}^{t_f} [x(s), Q(s) y(s)] ds$$

the optimal control u° can be shown to be satisfied by

$$(R + F^* Q F)u^\circ - (\hat{u} + F^* Q \hat{x}) = B^*(t) \Phi^*(t_0, t) c \quad (2.5.2-16)$$

or

$$u^\circ - \hat{u} = R^{-1} B^*(t) \lambda(t) \quad (2.5.2-17)$$

where $\lambda(t)$ is the solution to the following $2n$ linear differential equations

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\lambda}(t) \end{bmatrix} = \begin{bmatrix} A(t) & B(t) R^{-1}(t) B^*(t) \\ Q(t) & -A^*(t) \end{bmatrix} \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} + \begin{bmatrix} B(t) \hat{u}(t) \\ -Q(t)x(t) \end{bmatrix} \quad (2.5.2-18)$$

with the boundary conditions given by $x(t_0) = x_0$ and $x(t_f) = x_f$.

When $Q = R = I$, where I is the identity matrix, then equation (2.5.2-18) reduces to equation (2.5.2-8).

Example 2: Now let us consider $x(t_0) = x_0 \neq 0$ in Example 1 of Section 2.4.2, and for convenience let $Q = R = I$, where I = identity matrix.

For the present case

$$\hat{x} = x - \Phi(t, t_0)x_0 = Fu$$

we have

$$\begin{aligned} J(u) &= 1/2 \|x\|_{H_2}^2 + 1/2 \|u\|_{H_1}^2 \\ &= 1/2 \|Fu - \hat{y}\|_{H_2}^2 + 1/2 \|u\|_{H_1}^2 \end{aligned}$$

where

$$\hat{y} = -\Phi(t, t_0)x_0$$

Taking the advantage of the solution to Problem 4 (also see 2.5.2-5), the optimal u° is given by

$$(I + F^*F)u^\circ - F^*\hat{y} = B^*(t) \Phi^*(t_0, t)c$$

$$\text{or} \quad u^0 + F^*x = B^*(t)\Phi^*(t_0, t)c$$

$$\text{or} \quad u^0 = B^*(t)\Phi^*(t_0, t)\left[c - \int_t^t \Phi^*(s, t_0)x(s)ds\right]$$

By the same argument as stated in Section 3.1, the optimal $u(t)$ is given by

$$u(t) = B^*(t)\lambda(t)$$

where $\lambda(t)$ is the solution to the $2n$ linear differential equations given by (2.4.1-12).

Remark 1: In all previous examples the formulation for determining the optimal u for $x_0 = 0$ is also valid for $x_0 \neq 0$ by the application of projection $(\hat{u}, \hat{y})$ on the graph $H(F)$ following the method shown in Example 2 above.

Remark 2: It would be noted that the solution to Problem 4 in the first case is not $u^0 = \hat{u}$ because it does not belong to the set defined by

$$U_x = \left\{ x : u \in H, \quad F^t u = x \in \mathbb{R}^n \right\}$$

For free terminal condition, where the terminal point does not subject the constraint $F^t u = x$, the solution to Problem 4 is simply $u_0 = \hat{u}$ for the first case where $(\hat{x}, \hat{u}) \in H(F)$ and $u_0 = (I + F^*F)^{-1}(\hat{u} + F^*\hat{y})$ for the second case where $(\hat{x}, \hat{u}) \in H_1 \times H_2$.

Example 3: Let F^t be a random operator defined by

$$F^t = \sum_{i=1}^n \phi_i(t, a) > < f_i(t, a)$$

where

$$t \in \mathcal{T} = [t_0, t_f]$$

and $a \in \Omega = [\alpha, \beta]$, "a" is a random variable with known statistical properties.

Given $E(x) = \hat{x}$, where $E(x)$ = expectation of x , to find the element $u \in U$ such that

$$J(u) = E \int_{t_0}^{t_f} [(x - \hat{x}), (x - \hat{x})] ds + \int_{t_0}^{t_f} [u, u] ds \quad (2.5.2-19)$$

is minimized.

Assume:

(1) The initial point of x at $t=t_0$ is fixed, i.e., $x(t_0, a)$ is a fixed point.

(2) H_1 and H_2 spaces are defined by

$$H_1 = \left\{ u : \int_{t_0}^{t_f} [u, u] ds < \infty \right\}$$

$$H_2 = \left\{ x : \int_{t_0}^{t_f} E [x(s, a), x(s, a)] ds < \infty \right\}$$

respectively and the respective inner products on these spaces is given by the expression

$$\langle u, v \rangle = \int_{t_0}^{t_f} [u(s), v(s)] ds \quad u, v \in H_1$$

$$\langle x, y \rangle = \int_{t_0}^{t_f} E [x(s, a), y(s, a)] ds \quad x, y \in H_2$$

Then $J(u)$ in (2.5.2-19) may be written as

$$J(u) = \| x - \hat{x} \|_{H_2}^2 + \| u \|_{H_1}^2$$

and the present example is included within the original framework. This example can be solved by either of the following two methods.

First method: Because of randomness of "a" the terminal trajectory $x(t_f)$ at $t=t_f$ is not fixed. Also $\hat{u}=0$, $(\hat{x},0) \in$

$H_1 \times H_2$, and the relation $F\hat{x}=\hat{u}$ does not hold, hence the free terminal condition in the second case of Problem 4 will be considered. From the above definition of respective inner product on spaces H_1 and H_2 equations (AXI-1.2.5.1-10), and (2.5.1-11) become:

$$0 = \langle (x, y), (u, Fu) \rangle_{H(F)} = \langle x, u \rangle_{H_1} + \langle E(F^*y), u \rangle_{H_1} = \langle x + E(F^*y), u \rangle_{H_1} \quad (2.5.2-20)$$

$$(\hat{u}, \hat{y}) = (\bar{u}, F\bar{u}) + V(u_1, E(F^*u_1)) = (\bar{u} - E(F^*u_1), F\bar{u} + u_1) \quad (2.5.2-21)$$

and

$$\begin{aligned} \hat{u} &= \bar{u} - E(F^*u_1) \\ \hat{y} &= F\bar{u} + u_1 \end{aligned} \quad (2.5.2-22)$$

respectively.

Operating on the second equation with $E(F^*)$ of equation (2.5.1-22) and eliminating $E(F^*u)$ the result is

$$\bar{u} + E(F^*F\bar{u}) = \hat{u} + E(F^*\hat{y})$$

For the present example we have $\hat{u} = 0$, $\hat{x} = \hat{y}$, and $u = \bar{u}$, hence,

$$u + E(F^*Fu) = E(F^*\hat{x}) \quad (2.5.2-23)$$

Second method: $J(u)$ can also be written as

$$J(u) = E \langle x - \hat{x}, x - \hat{x} \rangle + \langle u, u \rangle$$

$$\text{where } \langle w, z \rangle = \int_{t_0}^t f[w(s), z(s)] ds$$

from the following chain of set equalities:

$$\begin{aligned} J(u) &= E \left\{ \langle x, x \rangle - 2 \langle \hat{x}, x \rangle + \langle \hat{x}, \hat{x} \rangle \right\} + \langle u, u \rangle \\ &= E \left\{ \langle Fu, Fu \rangle - 2 \langle \hat{x}, Fu \rangle + \langle \hat{x}, \hat{x} \rangle \right\} + \langle u, u \rangle \\ &= E \left\{ \langle F^*Fu, u \rangle - 2 \langle F^*\hat{x}, u \rangle + \langle \hat{x}, \hat{x} \rangle \right\} + \langle u, u \rangle \end{aligned} \quad (2.5.2-24)$$

the necessary condition to obtain the minimum value of $J(u)$ in a vector form is given by

$$\nabla_u J(u) = 0$$

$$\text{or} \quad 2 E(F^*Fu) - 2E(F^*\hat{x}) + 2u = 0$$

$$\text{or} \quad u + E(F^*Fu) = E(F^*\hat{x}) \quad (2.5.2-25)$$

Remarks: The optimal u in equation (2.5.2-23) (or equation 2.5.2-25) can be computed by the method of steepest descent for given $E(x) = \hat{x}$. Here we introduce the functional

$$f(u) = \langle \bar{R}u, u \rangle - 2 \langle u, \bar{g} \rangle \quad (2.5.2-26)$$

where $\bar{R}u = E(F^*F + I)u$ and $\bar{g} = E(F^*\hat{x})$. The problem of finding the optimal u in (2.5.2-23) (or 2.5.2-25) reduces to the problem of seeking the minimum of the functional in 2.5.2-26 (see Appendix AX).

CHAPTER III

COMPARISON WITH OTHER OPTIMIZING TECHNIQUES

It is interesting to make a comparison between several other methods of solutions for the problems stated in Chapter II. Here the case of continuous system in the example 1 of Problem 3 is considered for the purpose of this comparison.

3.1 Abstract Operator Formulation

It was proved in section 2.4.1 (see equation 2.4.1-8) that the element u which minimizes $J(u) = \|Fu\|_{H_2}^2 + \|u\|_{H_1}^2$ is given by $(I + F^*F)u = v$ for $v \in M$ and the solution of u to the above integral equation in operator form has been given by the lemma(4) stated in section 2.4.1. We shall prove this lemma in the following:

Since $v \in M$, it has a unique expansion in terms of any basis for that space. In the present instance the rows of matrix $\Phi(t_0, t)B(t)$ provide a convenient basis and introducing the scalars $c = \text{col.}(c_1, c_2 \dots c_n)$ we may write

$$v(t) = B^*(t) \Phi^*(t_0, t)c \quad (3.1-1)$$

as a defining relation for an arbitrary $v \in M$. Using the expression $x = Fu$ equation $(I + F^*F)u = v$ may be written

$$u = v - F^*x \quad (3.1-2)$$

It is not difficult to show (see Appendix AVI) that

$$(F^*x)(t) = B^*(t) \Phi^*(t_0, t) \int_t^{t_f} \Phi^*(s, t_0) x(s) ds \quad (3.1-3)$$

for $t \in \tau$

is the defining equation for F^* as t sweeps through τ .

Letting $\alpha(t)$ denote the tuplet

$$\begin{aligned} \alpha(t) &= c - \int_t^{t_f} \Phi^*(s, t_0) x(s) ds \\ &= c - \int_{t_0}^{t_f} \Phi^*(s, t_0) x(s) ds + \int_{t_0}^t \Phi^*(s, t_0) x(s) ds \\ &= \alpha(t_0) + \int_{t_0}^t \Phi^*(s, t_0) x(s) ds \end{aligned} \quad (3.1-4)$$

it is apparent that (3.1-2) may be written in the form

$$u(t) = B^*(t) \Phi^*(t_0, t) \alpha(t) \quad (3.1-5)$$

Consider now the vector $\lambda(t) = \Phi^*(t_0, t) \alpha(t)$, using the property that $\Phi^*(t_0, t) = \Psi(t, t_0)$ where $\Psi(t, t_0)$ is the transition matrix for the adjoint system

$$\dot{\Psi}(t, t_0) = -A^*(t) \Psi(t, t_0), \quad \Psi(t_0, t_0) = I \quad (3.1-6)$$

it is apparent from equation (3.1-4) that

$$\begin{aligned} \lambda(t) &= \Phi^*(t_0, t) \alpha(t_0) + \Phi^*(t_0, t) \int_{t_0}^t \Phi^*(s, t_0) x(s) ds \\ &= \Psi(t, t_0) \alpha(t_0) + \Psi(t, t_0) \int_{t_0}^t \Psi(t_0, s) x(s) ds \end{aligned} \quad (3.1-7)$$

which implies immediately that $\lambda(t)$ is the solution to the differential system.

$$\dot{\lambda}(t) = -A^*(t) \lambda(t) + x(t), \quad \lambda(t_0) = \lambda_0.$$

The original differential system for this example may now be written

$$\begin{aligned}\dot{x}(t) &= A(t)x(t) + B(t)u(t) \\ &= A(t)x(t) + B(t)B^*(t)\lambda(t)\end{aligned}$$

Thus in all we have the $2n$ linear differential equations

$$\begin{aligned}\dot{x}(t) &= A(t)x(t) + B(t)B^*(t)\lambda(t) & x(t_0) &= x_0 \\ \dot{\lambda}(t) &= -I x(t) - A^*(t)\lambda(t) & \lambda(t_0) &= \lambda_0\end{aligned}\quad (3.1-8)$$

the solution of which determines $\lambda(t)$, which in turn determines $u(t)$ by $u(t) = B^*(t)\lambda(t)$. (3.1-9)

One important point that must not be overlooked is that λ_0 is unknown. Indeed λ_0 is a function of c which we originally set out to determine. The solution of equations may be expressed in terms of its transition matrix $\hat{H}(t, t_0)$ which is defined by

$$\hat{H}(t, t_0) = \begin{bmatrix} A & BB^* \\ \hline I & -A^* \end{bmatrix} \hat{H}(t, t_0), \quad \hat{H}(t_0, t_0) = I \quad (3.1-10)$$

The transition matrix has $2n$ rows and $2n$ columns; it can be partitioned into four $n \times n$ submatrices

$$\hat{H}(t, t_0) = \begin{bmatrix} \hat{H}_{11}(t, t_0) & \hat{H}_{12}(t, t_0) \\ \hline \hat{H}_{21}(t, t_0) & \hat{H}_{22}(t, t_0) \end{bmatrix}$$

The solution to the system (3.1-8) thus can be expressed by

$$\begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} = \begin{bmatrix} \hat{H}_{11}(t, t_0) & \hat{H}_{12}(t, t_0) \\ \hat{H}_{21}(t, t_0) & \hat{H}_{22}(t, t_0) \end{bmatrix} \begin{bmatrix} x(t_0) \\ \lambda(t_0) \end{bmatrix} \quad (3.1-11)$$

In particular

$$x(t_f) = \textcircled{H}_{11}(t_f, t_0) x(t_0) + \textcircled{H}_{12}(t_f, t_0) \lambda(t_0) \quad (3.1-12)$$

Thus if $x(t_f)$ and $x(t_0)$ are known this expression determines $\lambda(t_0)$. In particular if $x(t_0) = 0$ and $\textcircled{H}_{12}(t, t_0)$ is non-singular then

$$\begin{aligned} \lambda(t_0) &= \textcircled{H}_{12}^{-1}(t_f, t_0) x(t_f) \\ \lambda(t) &= \textcircled{H}_{22}(t, t_0) \textcircled{H}_{12}^{-1}(t_f, t_0) x(t_f) \end{aligned} \quad (3.1-13)$$

and finally the control u is given by

$$u(t) = B^*(t) \textcircled{H}_{22}(t, t_0) \textcircled{H}_{12}^{-1}(t_f, t_0) x(t_f) \quad (3.1-14)$$

3.2 Dynamic Programming Approach: (10), (45), (75)

If the dynamic plant and the index of performance are respectively by

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

$$J = T_e(x_f, t_f) + \int_{t_0}^{t_f} C(\varphi_u(s, x_0, t_0), u(s), s) ds$$

it can be shown that a dynamic programming approach leads to the equation

$$-\frac{\partial J^0}{\partial t_1} = \min_{u \in U} \left[C + \sum_{i=1}^n \frac{\partial J^0}{\partial x_i} \dot{x}_i \right]_{t=t_1} \quad (3.2-1)$$

where J^0 is the unknown optimal value of the index of performance. To consider a general initial time, we replace t_1 by t and rewrite (3.2-1) in the following form

$$-\frac{\partial J^\circ}{\partial t} = \min_{u \in U} \left\{ C + [\nabla_x J^\circ, f(u, x, t)] \right\} \quad (3.2-2)$$

where $\nabla_x J^\circ = \text{grad}_x J^\circ$

Let us now define a function

$$H(x, \nabla_x J^\circ, t) = \min_{u \in U} \left\{ C + [\nabla_x J^\circ, f(u, x, t)] \right\}$$

then J° satisfies the Hamilton-Jacobi equation

$$\frac{\partial J^\circ}{\partial t} + H(x, \nabla_x J^\circ, t) = 0 \quad (3.2-3)$$

with the boundary $J(x, t) = T_e(x, t)$ for x and t on the terminal set S . For the minimum energy problem $T_e(x, t) = 0$, the boundary condition in solving J° becomes

$$J(x_f, t_f) = 0 \quad \text{for } x_f \text{ and } t_f \text{ on the terminal set } S.$$

In general (3.2-3) is very difficult to solve. However, if the solution of J° in (3.2-3) has the quadratic form in x with a time varying matrix coefficient $P(t)$, and $P(t)$ is symmetric then

$$J^\circ = \frac{1}{2} [x, P(t)x], \quad (3.2-4)$$

where $P(t_f) = 0$ in order that $J(x_f, t_f) = 0$ is satisfied.

Taking partial derivatives of (3.2-4) gives

$$\frac{\partial J^\circ}{\partial t} = \frac{1}{2} [x, \dot{P}(t)x] \quad (3.2-5)$$

$$\text{and} \quad \nabla_x J^\circ = P(t)x \quad (3.2-6)$$

Carrying out the minimization of (3.2-2) the optimal control function $u(t)$ can be expressed by

$$u = -R^{-1}B (\nabla_x J^\circ) = -R^{-1}B^* P(t)x \quad (3.2-7)$$

where $C = \frac{1}{2} \{ [x, Qx] + [u, Ru] \}$ and $f = Ax + Bu$.

Substituting (3.2-4), (3.2-5) and (3.2-6) in (3.2-2) (or 3.2-3) and taking twice the partial derivatives with respect to x the final result is given by

$$-\frac{dP(t)}{dt} = Q + PA + A^*P - PBR^{-1}B^*P \quad (3.2-8)$$

Equation (3.2-8) is known as the matrix Riccati equation with the boundary condition $P(t_f) = 0$, and in general is a nonlinear differential equation which only in special case can be solved in closed form. Once $P(t)$ is solved by (3.2-8) the optimal u^0 and J^0 can be computed. Equation (3.2-7) shows that the optimal control law is a linear function of x in which all state variables of the system must be known at all times to effect control. The matrix $K(t) = R^{-1}(t)B^*(t)P(t)$ will be called the optimal gain.

3.3 Pontryagin's Approach: (21), (68)

In applying Pontryagin's maximum principle, an auxiliary vector $\psi(t) = (\psi_1(t) \dots \psi_n(t))$ is defined and the function $H(\psi, x, u)$ must be found such that

$$H(\psi, x, u) = -C + [\psi, \dot{x}] \quad (3.3-1)$$

where for the Example 1 of Problem 3 in Section (2.4.2)

$$C = \frac{1}{2} \{ [x, Qx] + [u, Ru] \}$$

$$\dot{x} = Ax + Bu$$

The maximum principle states that in order to minimize the index of performance J the scalar function $H(\psi, x, u)$ has to

be maximized with respect to all the components of u at any time t . Since the control signal $u(t)$ is unconstrained in this example, then the necessary condition for a minimum of H with respect to u written in vector form is

$$\nabla_u H = \text{col.} \left[\frac{\partial H}{\partial u_1} \dots \frac{\partial H}{\partial u_m} \right] = 0 \quad (3.3-2)$$

From (3.3-1) and (3.3-2) the optimal control function u° for the Example 1 of Problem 3 is obtained

$$u^\circ = R^{-1} B^* \psi \quad (3.3-3)$$

Combining (3.3-3) with the following Hamiltonian system

$$\begin{aligned} \frac{dx_1}{dt} &= \frac{\partial H}{\partial \psi_1} & 1 &= 1 \dots n \\ \frac{d\psi_1}{dt} &= -\frac{\partial H}{\partial x_1} & 1 &= 1 \dots n \end{aligned} \quad (3.3-4)$$

gives

$$\begin{aligned} \dot{x} &= Ax + Bu \\ \dot{\psi} &= -A^* \psi + Qx \end{aligned}$$

or

$$\begin{bmatrix} \dot{x} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} A & BR^{-1}B^* \\ Q & -A^* \end{bmatrix} \begin{bmatrix} x \\ \psi \end{bmatrix} \quad (3.3-5)$$

For the boundary conditions $x(t_0) = 0$ and $x(t_f) = x_f$, the solution of $u(t)$ is given by (2.4.2-12).

The vector $\psi(t)$ can be interpreted geometrically as follows: Let

$$\psi(t) = -P(t)x(t) \quad (3.3-6)$$

and take the total derivatives with respect to time of equation

(3.3-6), it gives

$$\dot{\psi} = -[\dot{P}x + P(Ax + Bu)] \quad (3.3-7)$$

From (3.3-3) and (3.3-5) equation (3.3-7) reduces to

$$Qx + A^*Px = -\dot{P}x - PAx + PBB^{-1}B^*Px \quad (3.3-8)$$

Taking partial derivatives with respect to x of (3.3-8), it gives

$$-\dot{P} = Q + PA + A^*P - PBB^{-1}B^*P \quad (3.3-9)$$

which shows that $P(t)$ in (3.3-6) satisfies the matrix Riccati equation. Comparison of (3.2-6) with (3.3-6) results in $\psi(t) = -\nabla_x J^0$. Hence $\psi(t)$ represents a vector that points in the direction where J^0 decreases most rapidly.

3.4 Calculus of Variations (13), (28)

$$\text{To minimize the integral } J = \frac{1}{2} \int_{t_0}^{t_f} \left\{ [x, Qx] + [u, Ru] \right\} ds \quad (3.4-1)$$

with the constraint $g = \dot{x} - Ax - Bu = 0$

we construct the function F such that

$$\begin{aligned} F &= C + [\lambda, g] \\ &= \frac{1}{2} [x, Qx] + \frac{1}{2} [u, Ru] + [\lambda, \dot{x} - Ax - Bu] \end{aligned} \quad (3.4-2)$$

where λ is an unknown time-dependent Lagrangian multiplier vector. The conditions to minimize the performance index J are given by the following vector equations

$$\nabla_x F - \frac{d}{dt} \nabla_{\dot{x}} F = 0$$

$$\nabla_u F - \frac{d}{dt} \nabla_{\dot{u}} F = 0$$

(3.4-3)

$$\nabla_{\lambda} F - \frac{d}{dt} \nabla_{\dot{\lambda}} F = 0$$

which gives

$$\dot{\lambda} = Qx - A^* \lambda$$

$$u = R^{-1} B^* \lambda$$

(3.4-4)

$$\dot{x} = Ax + Bu$$

From (3.4-4) it follows that

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} A & BR^{-1}B^* \\ Q & -A^* \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix}$$

(3.4-5)

(3.4-5) gives results identical to those obtained for the system given by (3.3-5) and therefore the solution for optimal $u(t)$ has the same form as (3.3-3).

3.5 A Method for Evaluating the Optimal Performance Index

Given by

$$J^0 = \frac{1}{2} \int_{t_0}^{t_f} \left\{ [x(s), Q(s)x(s)] + [\dot{u}(s), R(s)\dot{u}(s)] \right\} ds \quad (3.5-1)$$

The evaluation of the optimal performance index J^0 given by (3.5-1) is a difficult task. A compact form of $J^0(u)$ which is easily computed will be derived in this section. It will be found to be useful in the proof of the equivalent expression

for the optimal control function $u^{\circ}(t)$, the optimal performance index J° derived from the functional analysis, and other techniques.

It was found in the previous section that the optimal control function $u^{\circ}(t)$ has the form

$$u^{\circ}(t) = -R^{-1}(t)B^*(t)P(t)x(t)$$

where $P(t)$ is the solution of the matrix Riccati equation

$$-\dot{P}(t) = Q + PA + P^*A - PBR^{-1}B^*P$$

with the boundary condition given by $P = S$, which depends on the terminal condition of $x(t)$ and the nature of the problem under consideration. Substituting the values of Q and u° in the expression of J° , the optimal performance index J° has the form

$$J^{\circ} = \frac{1}{2} \int_{t_0}^{t_f} \left\{ [x, (-\dot{P} - PA - A^*P + PBR^{-1}B^*P)x] + [R^{-1}B^*Px, B^*Px] \right\} dt \quad (3.5-2)$$

where x in (3.5-2) is the optimal state variable. Because of the following properties

$$[R^{-1}B^*Px, B^*Px] = [x, P^*BR^{-1}B^*Px]$$

$$PA = \frac{1}{2} (PA + A^*P)$$

Where R is symmetric positive definite matrix and P is symmetric, and in the optimal trajectory $\dot{x} = Ax - BR^{-1}B^*Px$ equation (3.5-2) is simplified to the following form

$$J^{\circ} = -\frac{1}{2} \int_{t_0}^{t_f} \left\{ [x, (\dot{P}x + 2P\dot{x})] \right\} dt \quad (3.5-3)$$

$$\text{or } J^0 = -\frac{1}{2} \int_{t_0}^{t_f} [x, (\dot{P}x + P\dot{x})] dt - \frac{1}{2} \int_{t_0}^{t_f} [x, P\dot{x}] dt \quad (3.5-4)$$

Integrating by parts the first term in (3.5-4) gives

$$\int_{t_0}^{t_f} [x, (\dot{P}x + P\dot{x})] dt = [x, Px]_{t_0}^{t_f} - \int_{t_0}^{t_f} [Px, \dot{x}] dt \quad (3.5-5)$$

Substituting (3.5-5) into (3.5-4), the final result of the optimal J^0 has the following form:

$$J^0 = \frac{1}{2} [x(t_0)p(t_0)x(t_0)] - \frac{1}{2} [x(t_f), p(t_f)x(t_f)] \quad (3.5-6)$$

or

$$J^0 = \frac{1}{2} [x(t_f), \psi(t_f)] - \frac{1}{2} [x(t_0), \psi(t_0)] = \frac{1}{2} [x(t), \psi(t)]_{t_0}^{t_f} \quad (3.5-7)$$

where $\psi(t)$ is the solution given by (3.3-5).

A generalized Kalman's method for the determination of optimal performance index has been developed and given in (3.5-7). It can be applied to the fixed boundary problems as well as the variable boundary problems. We thus have the following lemma:

lemma: The optimal performance index J^0 for the dynamic plant described by $\dot{x} = A(t)x(t) + B(t)u(t)$, with $x(t_0) = x_0$, $x(t_f) = x_f$, and the performance index

$$J = \frac{1}{2} \int_{t_0}^{t_f} \{ [x, Qx] + [u, Ru] \} dt, \text{ where } Q \text{ and } R \text{ are symmetric}$$

positive definite matrices, has the form

$$J^0 = \frac{1}{2} [x(t), \psi(t)]_{t_0}^{t_f}$$

where $\psi(t)$ is the solution given by (3.3-5)

Example 1: The optimal performance index of Example 1 in Section 2.2.2 is computed by

$$J^0 = \frac{1}{2} [x(t_f), \psi(t_f)] - \frac{1}{2} [x(t_0), \psi(t_0)] \quad (3.5-8)$$

where

$$\psi(t_f) = \mathbb{H}_{21}(t_f, t_0)x(t_0) + \mathbb{H}_{22}(t_f, t_0)\psi(t_0),$$

$$\psi(t_0) = \mathbb{H}_{12}(t_f, t_0)^{-1} [x(t_f) - \mathbb{H}_{11}(t_f, t_0)x_0],$$

and the grand matrix $\mathbb{H}(t, t_0)$ is defined by

$$\dot{\mathbb{H}} = \begin{bmatrix} A & B B^* \\ 0 & -A^* \end{bmatrix} \mathbb{H}(t, t_0), \quad (3.5-9)$$

$$\mathbb{H}(t_0, t_0) = I$$

Example 2: The optimal performance index of Example 1 in Section 2.4.2 is given by

$$J^0 = \frac{1}{2} [x(t_f), \psi(t_f)] \quad (3.5-10)$$

where $\psi(t_f) = \mathbb{H}_{22}(t_f, t_0)\psi(t_0)$

$$\psi(t_0) = \mathbb{H}_{12}^{-1}(t_f, t_0)x(t_f)$$

the grand matrix $\mathbb{H}(t, t_0)$ is defined by

$$\dot{\mathbb{H}} = \begin{bmatrix} A & B R^{-1} B^* \\ Q & -A^* \end{bmatrix} \mathbb{H}(t, t_0), \quad \mathbb{H}(t_0, t_0) = I \quad (3.5-11)$$

3.6 The Equivalent Expression of Optimal Control Function and Optimal Performance Index

3.6.1 Example 1. Proof of Equivalence for Example 1 of Problem 1 in Section 2.2.2

We shall prove the equivalence of

$$u^{\circ} = \sum_{i=1}^n f_i \alpha_i \quad (3.6.1-1)$$

$$J^{\circ} = \frac{1}{2} \langle \beta, \mathcal{F}^{-1} \beta \rangle \quad (3.6.1-2)$$

$$\text{and } u^{\circ} = B^*(t) \lambda(t) \quad (3.6.1-3)$$

$$J^{\circ} = \frac{1}{2} [x(t_f), \lambda(t_f)] - \frac{1}{2} [x(t_0), \lambda(t_0)] \quad (3.6.1-4)$$

respectively, where $\lambda(t)$ is the solution of (3.4-5) with $R = I$, $Q = 0$. The abstract operator form of (3.6.1-1) may be written in terms of matrix expression

$$u^{\circ} = B^*(t) \Phi^*(t_0, t) \mathcal{F}^{-1} \beta \quad (3.6.1-5)$$

$$\text{where } \beta = \Phi(t_0, t_f) x_f - x_0 \quad (3.6.1-6)$$

$$\alpha = \mathcal{F}^{-1} \beta = \mathcal{F}^{-1} [(t_0, t_f) x_f - x_0] \quad (3.6.1-7)$$

$$\text{and } \mathcal{F} = \int_{t_0}^{t_f} \Phi(t_0, s) B(s) R^{-1}(s) B^*(s) \Phi^*(t_0, s) ds$$

$$\text{Let } \lambda(t) = \Phi^*(t_0, t) \mathcal{F}^{-1} \beta = \Phi^*(t_0, t) \alpha \quad (3.6.1-8)$$

then (3.6.1-5) has the same form as (3.6.1-3).

It is sufficient to prove that $\lambda(t)$ in (3.6.1-8) satisfies the system of $2n$ equations given by (2.4.1-12)

(or 3.4-5 with $R = I$, $Q = 0$)

Since $\Phi^*(t_0, t_0) = I$

equation (3.6.1-8) may be written as

$$\lambda(t) = \Phi^*(t_0, t)\lambda(t_0) = \Phi^*(t_0, t)\lambda_0 \quad (3.6.1-9)$$

$$\text{here } \alpha = \lambda_0 \quad (3.6.1-10)$$

Solve x_f in (3.6.1-7) and noting that $\Phi(t_0, t_f)^{-1} = \Phi(t_f, t_0)$

gives $x(t_f) = \Phi(t_f, t_0)x_0 + \Phi(t_f, t_0)$

$$\int_{t_0}^{t_f} \Phi(t_0, s)B(s)B^*(s)\Phi^*(t_0, s)\lambda_0 ds \quad (3.6.1-11)$$

$$\text{or } x(t_f) = \Phi(t_f, t_0)x_0 + \Phi(t_f, t_0) \int_{t_0}^{t_f} \Phi(t_0, s)B(s)B^*(s)\lambda(s)ds$$

Since it holds for all t_f , we may replace t_f by a general time t

$$x(t) = \Phi(t, t_0)x_0 + \Phi(t, t_0) \int_{t_0}^t \Phi(t_0, s)B(s)B^*(s)\lambda(s)ds \quad (3.6.1-12)$$

where $\Phi(t, t_0)$ is the transition matrix which satisfies the homogeneous system

$$\dot{x}(t) = A(t)x(t)$$

$$x(t_0) = x_0 \quad (3.6.1-13)$$

Let $\Psi(t, t_0)$ be the transition matrix that satisfies the adjoint system of (3.6.1-13) that is

$$\dot{\lambda}(t) = -A^*(t)\lambda(t)$$

$$\lambda(t_0) = \lambda_0 \quad (3.6.1-14)$$

Now since the transition matrices Φ and Ψ have the following property

$$\Psi(t, t_0) \Phi^*(t, t_0) = I \quad (3.6.1-15)$$

equation (3.6.1-9) becomes

$$\lambda(t) = \Psi(t, t_0) \lambda_0 \quad (3.6.1-16)$$

Equation (3.6.1-12) and (3.6.1-16) implies that $x(t)$ and $\lambda(t)$ are satisfied by the following linear system of $2n$ equations

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} A & BB^* \\ 0 & -A^* \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} \quad (3.6.1-17)$$

with $x(t_0) = x_0$, $x(t_f) = x_f$

Next we shall prove the equivalence of (3.6.1-2) and (3.6.1-4).

In view of (3.6.1-6), (3.6.1-7) and (3.6.1-10) equation

(3.6.1-2) reduces to

$$\begin{aligned} J &= \frac{1}{2} \langle \beta, \mathcal{F}^{-1} \beta \rangle \\ &= \frac{1}{2} [\Phi(t_0, t_f) x_f - x_0, \lambda_0] \\ &= \frac{1}{2} [\Phi(t_0, t_f) x_f, \lambda_0] - \frac{1}{2} [x_0, \lambda_0] \\ &= \frac{1}{2} [x_f, \Phi^*(t_0, t_f) \lambda_0] - \frac{1}{2} [x_0, \lambda_0] \\ &= \frac{1}{2} [x(t_f), \lambda(t_f)] - \frac{1}{2} [x(t_0), \lambda(t_0)] \end{aligned}$$

where $\lambda(t)$ is the solution of (2.4.1-12).

3.6.2 Example 2. Proof of Equivalence for the Example 2 in Section 2.3.2

Let us solve the example of Problem 2 in Section 2.3.2 by the method of calculus of variations.

Given the differential equation: $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ with boundary conditions $x(t_f) = 0$ and $\psi = [x(t_0), x(t_0)] - 1 = 0$,
 and performance index $J = \int_{t_0}^{t_f} [u, u] ds$, to

find $u(t)$ such that J takes on the least possible value.

From (3.4-4) the solution of x at $t=t_f$ is given by

$$0 = x(t_f) = \Phi(t_f, t_0) \left[x_0 + \int_{t_0}^{t_f} \Phi(t_0, s) B(s) \right.$$

$$\left. B^*(s) \Phi^*(t_0, s) ds \lambda_0 \right]$$

$$\text{hence } x_0 = -\mathcal{J} \lambda_0 \quad (3.6.2-2)$$

To determine λ_0 we must turn to the transversality conditions. Since t_0 , t_f and $x_1(t_f)$ are fixed all derivatives with respect to these variables are zero.

The transversality condition

$$\left(F - \sum_{i=1}^n \dot{x}_i F_{\dot{x}_i} \right) \bigg|_{t_0} \delta t_0 + \sum_{i=1}^n F_{\dot{x}_i} \delta x_i \bigg|_{t_0}^{t_f} + \sum_{k=1}^p \eta_k \delta \psi_k \bigg|_{t_0} = 0.$$

reduces to

$$- \sum_{i=1}^n F_{\dot{x}_i} \delta x_i \bigg|_{t_0} + \eta_1 \frac{\partial \psi_1}{\partial x_1} \delta x_1 \bigg|_{t_0} = 0 \quad (3.6.2-3)$$

Since $F = [u, u] + [\lambda, \dot{x} - Ax - Bu]$, (3.6.2-3) becomes

$$-\lambda_1(t_0) + \eta_1 \frac{\partial \psi_1}{\partial x_1} = 0 \quad (3.6.2-4)$$

In vector form

$$-\lambda(t_0) + \eta_1 2x_0 = 0$$

$$\text{or } x_0 = \frac{1}{2\eta_1} \lambda_0 \quad (3.6.2-5)$$

from (3.6.2-1) and (3.6.2-5) we have

$$[x_0, x_0] = \frac{1}{4\eta_1^2} [\lambda_0, \lambda_0] = 1$$

$$\text{and therefore } \eta_1 = \pm \frac{\|\lambda_0\|}{2}$$

$$\text{or } \lambda_0 = \pm \|\lambda_0\| x_0 \quad (3.6.2-6)$$

Substituting (3.6.2-6) in (3.6.2-2) we have

$$x_0 = \mp \mathcal{F} \|\lambda_0\| x_0 \quad (3.6.2-7)$$

But the optimal performance index is given by

$$J^0 = \langle x(t), \lambda(t) \rangle \Big|_{t_0}^{t_f} \quad (3.6.2-8)$$

when $x(t_f) = 0$ (3.6.2-8) becomes

$$J^0 = -\langle x_0, \lambda_0 \rangle \quad (3.6.2-9)$$

Substituting (3.6.2-6) in (3.6.2-9) it reduces to

$$J^0 = \mp \langle x_0, \|\lambda_0\| x_0 \rangle = \mp \|\lambda_0\| \langle x_0, x_0 \rangle = \mp \|\lambda_0\|$$

Since $J^0 \geq 0$ so we take the positive sign.

Hence

$$J^0 = \|\lambda_0\|$$

which gives

$$\mathcal{F}^{-1} x_0 = J^0 x_0 \quad (3.6.2-10)$$

In terms of the notation $\beta = -x_0$, as expressed in the functional analysis approach we have

$$\mathcal{F}^{-1}\beta = J^{\circ}(\beta)\beta$$

TABLE 2
COMPARISON OF OPTIMIZATION TECHNIQUES

Example 1 in Section 2.4.2	Dynamic Programming	Maximum principle	Calculus of variations	Functional analysis
Example 1 in Section 2.2.2	Solving nonlinear Riccati equation $\dot{P} = PA + A^*P - PBB^*P$ with $P(t_f) = 0$	Solving $2n \times 2n$ linear differential equation given by 3.6.1-17	Solving $2n \times 2n$ linear differential equation given by 3.6.1-17	Determination of the coefficient d_1
	optimal u^0 $u^0 = -B^*Px$	$u^0 = B^*(t) \left\{ \begin{matrix} \textcircled{H}_{21}(t, t_0)x_0 \\ + \textcircled{H}_{22}(t, t_0) \textcircled{H}_{12}^{-1}(t_f, t_0) \\ [x_f - \textcircled{H}_{11}(t_f, t_0)x_0] \end{matrix} \right\}$ where $\textcircled{H}$ is defined by 3.5-9	Same result as obtained from the maximum principle	$u^0 = \sum_{i=1}^n f_i d_i$
Example 1 in Section 2.4.2	Solving nonlinear Riccati equation $\dot{P} = PA + A^*P + Q - PBR^{-1}$ B^*P with $P(t_f) = 0$	Solving $2n \times 2n$ linear differential equation given by 3.3-5	Same as the maximum principle	Determination of the coefficient C_1 that satisfies $(R + P^*QP)u = \sum_{i=1}^n f_i C_i$
	optimal u^0 $u^0 = R^{-1}B^*Px$	$u^0 = R^{-1}(t)B^*(t) \left\{ \begin{matrix} \textcircled{H}_{21}^{-1}(t, t_0) \textcircled{H}_{12} \\ (t_f, t_0)x_f \end{matrix} \right\}$ Where $\textcircled{H}$ is defined by 3.5-11	Same as the maximum principle	Same as the maximum principle (or calculus of variations)

CHAPTER IV

CONCLUSIONS AND SUGGESTIONS FOR FUTURE EXTENSION

The contribution of this dissertation are reviewed in this chapter. Certain aspects worthy of future investigation are discussed.

4.1 Conclusions:

In this thesis, abstract operator methods are applied to a class of linear optimal control problems. The main theories related to the optimization problems have been developed and several examples of optimal trajectory are given. The method and solution to these optimal control problems presented are concise in form and concept, and very general in scope.

Four types of minimum energy problems have been studied in Chapter II. Problem 1 deals with the minimum energy problem of finding a u with minimum norm such that $F^t u = x$ where the linear and bounded operator F which maps the Hilbert function space U onto R^n , is defined by

$$F^t = \sum_{i=1}^n \phi_i \rangle \langle f_i \quad (4.1-1)$$

and $\{\phi_i\}_1^n$ and $\{f_i\}_1^n$ are linearly independent.

In Problem 2, a parametrized family $\{F^s : s \in \mathcal{V}\}$ of linear transforms each of which is a mapping of the Hilbert

space U onto R^n has been considered. A generalization of minimum energy problems has been developed and is discussed in Problem 3 and 4. By the concept of graph of transformation the elements $u \in U$ and $x \in X$ with minimum norm given by

$$J(u) = \|Fu\|_{H_2}^2 + \|u\|_{H_1}^2 \quad (\text{Note } x = Fu) \text{ can be determined.}$$

From the generalization of Kalman's method, the desired minimum energy (i.e., the optimal performance index of the generalized minimum energy problem) can be expressed by equation (3.5-7). The class of problems with which we have been concerned in Chapter II were centered about a transformation of Hilbert function space H onto R^n . The basic principle of seeking the element in H with the minimum norm is to locate the orthogonal complement of the null space of the respective operator under consideration. This method is attractive in that it is applicable to continuous or discrete linear dynamic systems for both fixed and variable boundary conditions.

The related optimal parameters are determined by the method of steepest descent in the Hilbert space. The technique of formulating the iterative equation and the convergence of the sequence have been discussed in detail in the Appendix AX.

Application of Problem 3 to the study of the open loop system reliability, has been studied in Example 3 in Section 2.4.2.

This dissertation has discussed only a class of linear control problem, namely, minimum energy problems which are applicable to the solution of continuous (or discrete) linear

dynamic system with no constraint on $u(t)$.

Within this dissertation no specific mention was made regarding the stability of solution of the system defined by equation (1.1.2).

When t_f is allowed to become infinite, the stability of the optimal system of which the performance index is given by

$$J = \frac{1}{2} \int_{t_0}^{t_f} \{ [x, Qx] + [u, Ru] \} dt \quad (4.1-2)$$

can be investigated in the following.

The optimal control function $u(t)$ can be expressed by (see 3.2-7)

$$u = R^{-1}B^*Px \quad (4.1-3)$$

where $P(t)$ is symmetric.

Substituting (4.1-3) in (4.1-2), considering the optimal value of the performance index in (4.1-2) and replacing t_0 by t , the following function is obtained

$$\begin{aligned} J^0 &= \int_t^{\infty} \{ [x, Qx] + [R^{-1}B^*Px, B^*Px] \} dt \\ &= \int_t^{\infty} [x, (Q + PBR^{-1}B^*P)x] dt \end{aligned}$$

Since the matrices Q and R are positive definite, J^0 is positive for any initial value $x(t)$ at any time t . From the theory of dynamic programming, the total derivative J^0 with respect to time is found to be

$$\frac{dJ^0}{dt} = - [x, (Q + PBR^{-1}B^*P)x] < 0$$

Therefore J^0 can be used as a Liapunov function to prove that the system described by (1.1.2) with J given by (4.1-2) is asymptotically stable. (21,53)

4.2 Suggestions for Future Work

Several problems of an associated nature are not covered by the present thesis. The six problem areas listed below are indicative of the potential impact of the function space approach on control technology. Each case is known to be an area of current research activity and in some instances much progress has already been made.

- (1) Consideration of minimum energy problems discussed in this dissertation with constraint on the control function $u(t)$.
- (2) Generalization of the minimum energy problem to the composite dynamic linear system (a combination of continuous and discrete systems) with an admissible performance index given by

$$J(u) = \sum_{\alpha=0}^{[t_f]-1} h_{\alpha} [u(t_{\alpha}), R(t_{\alpha})u(t_{\alpha})] + \int_{t_0}^{t_f} [(x-\hat{x}), Q(x-\hat{x})] ds \quad (4.2-1)$$

where Q and R are positive definite matrices.

- (3) Extension of abstract operator formulation to the solution of minimum effort control problem. Here we may consider the operator F as a transformation defined on a Banach space and having range R^n . If we

treat the control function $u(t)$, $t_0 \leq t \leq t_f$ as an element in a Banach space, say $L_p(t_0, t_f)$, then the performance index defined by

$$J(u) = \left[\int_{t_0}^{t_f} |u(t)|^p dt \right]^{\frac{1}{p}} \quad (4.2-2)$$

is exactly the norm of $u(t) \in L_p(t_0, t_f)$. In other words, the input space $L_p(t_0, t_f)$ is a Banach space for which $J(u)$ is the norm of $u(t)$. The minimum effort problem may be formulated in the following:

Given arbitrary $x \in R^n$ find $u \in B$ with minimum norm such that $Fu = x$

- (4) Application of functional analysis to solve a class of nonlinear system described by

$$\dot{x}(t) = f(x, t) + g(u, t) \quad (4.2-3)$$

- (5) Extension to an adaptive optimum system such as the one described by the following block diagram.

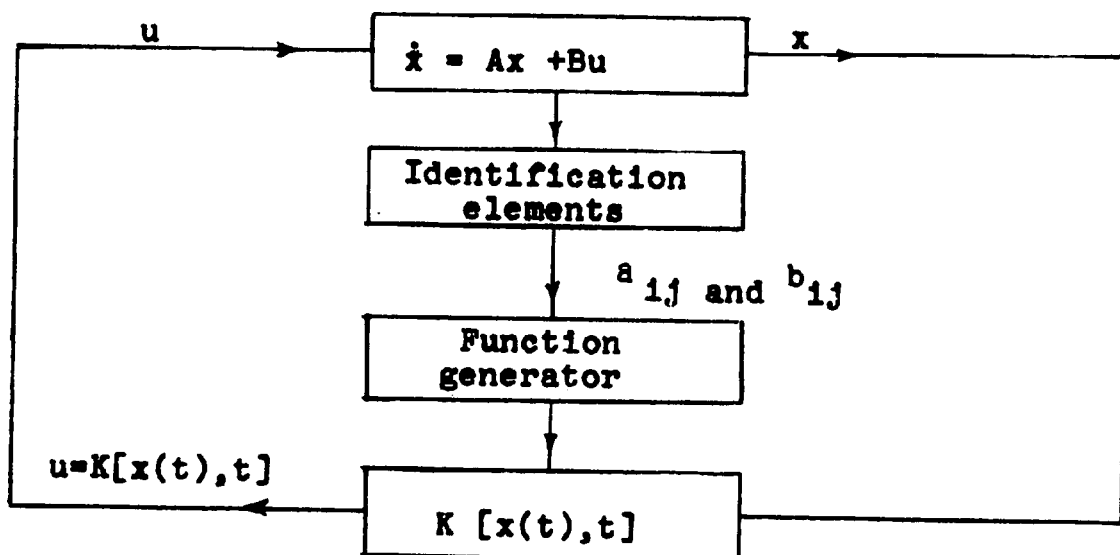


Fig. 4.2-1.--Adaptive optimum system.

- (6) Development of abstract operator formulation to the solution of minimum energy problem when the system is subjected to a random input disturbance, such as the system whose dynamic behavior is given by

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + \eta(t) \quad (4.2-4)$$

where $\eta(t)$ is a random input disturbance.

APPENDIX AI

REVIEW OF BASIC TERMINOLOGY IN FUNCTIONAL ANALYSIS

AI-1 Metric Space:

A set X is called a metric space if to every pair of elements x, y . (also called points or vectors) of this set there corresponds a non negative real number $\rho(x, y)$ such that

- (1) $\rho(x, y) = 0$ if and only if $x = y$ (identity)
- (2) $\rho(x, y) = \rho(y, x)$ (symmetry)
- (3) $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$ (triangle inequality)

The number $\rho(x, y)$ is called the metric function for the space and can be thought of as an abstract distance between the elements x and y . The above three conditions are called the metric axioms.

- (1) The space $L_p(a, b)$

$$L_p = \left\{ x(t) : \int_a^b |x(t)|^p dt < \infty \right\}$$

If $x(t) \in L_p$ and $y(t) \in L_p$, we define

$$\rho(x, y) = \left[\int_a^b |x(t) - y(t)|^p dt \right]^{\frac{1}{p}}$$

we can prove that the metric axioms are fulfilled.

- (2) The space l_p

$$l_p = \left\{ \{ \xi_i \} : \sum_{i=1}^{\infty} |\xi_i|^p < \infty \right\}$$

If $x(t) = \{\xi_1 \dots \xi_n \dots\} \in \ell_p$ and $y(t) = \{\eta_1 \dots \eta_n \dots\} \in \ell_p$

we define the metric distance by

$$\rho(x, y) = \left(\sum_{i=1}^{\infty} |\xi_i - \eta_i|^p \right)^{\frac{1}{p}}$$

we see at once that the metric axioms hold. It should be noted that the matrix space is not required to be complete in the above examples.

AI-2 Linear Space:

A set X of elements $x, y, \dots$ is called a linear space if addition and multiplication are defined on X and if the following rules hold:

- 1 (a) $x + y = y + x$ for x and y in X
- (b) $(x + y) + z = x + (y + z)$ for any x, y, z in X
- (c) There exists an element $0 \in X$ (the "zero element") such that

$$x + 0 = x \text{ for any } x \in X$$
- (d) for any $x \in X$ there exists an element $y \in X$ (the inverse element) such that $x + y = 0$
- 2 (e) $1 \cdot x = x$ for any $x \in X$
- (f) $\alpha(\beta x) = (\alpha\beta)x$ for any $x \in X$ and any scalar α and β
- 3 (g) $(\alpha + \beta)x = \alpha x + \beta x$ for any $x \in X$ and any α and β
- (h) $\alpha(x + y) = \alpha x + \alpha y$ for any $x, y \in X$ and any α

A linear space is called "real" or complex according to whether the scalars $(\alpha, \beta, \dots)$ are in the real or

complex number system.

Example $C(a,b)$, $L_p(a,b)$ and ℓ_p are linear spaces.

AI-3 Norm Linear Space:

A set X is called the norm space if to every vector x there corresponds a real number, denoted by $\|x\|$ and which is called the norm of x . It possesses the following properties:

- (1) $\|x\| \geq 0$ and $\|x\| = 0$ only if $x = 0$
- (2) $\|x + y\| \leq \|x\| + \|y\|$
- (3) $\|\alpha x\| = |\alpha| \|x\|$

In a normed linear space we can introduce a metric by

$$\rho(x,y) = \|x - y\|$$

The convergence in a normed linear space (convergence in the norm) is defined by $\|x_n - x\| \rightarrow 0$ as $n \rightarrow \infty$.

AI-4 Banach Space:

If every fundamental sequence (Cauchy sequence) of a metric space X has a limit in X then X is called complete.

A Banach space is a complete normed linear space.

Example 1: $C(a,b)$, $L_p(a,b)$ and ℓ_p are Banach spaces where the norms of these spaces are respectively defined by

$$\|x\| = \max_{a \leq t \leq b} |x(t)|, \quad x \in C(a,b)$$

$$\|x\| = \left(\int_a^b |x(t)|^p dt \right)^{\frac{1}{p}}, \quad x \in L_p(a,b)$$

$$\|x\| = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{\frac{1}{p}}, \quad x = (x_1, \dots, x_n, \dots) \in \ell_p$$

Example 2: The n dimensional vector space E_n is a Banach space where the norm is defined by

$$\|x\| = \left(\sum_{i=1}^n |\xi_i|^2 \right)^{\frac{1}{2}}, \quad x = (\xi_1 \dots \xi_n) \in E_n$$

AI-5 Inner Product Space:

If in a normed linear space, the length, or norm, of a vector is defined in term of a certain of two vectors this space is called an inner product space.

AI-6 Hilbert Space:

An inner product metric space, which is infinite dimensional and complete, is called a Hilbert space (real or complex). If X is a finite-dimensional inner product space, then it is called a Euclidean space.

Example 1: The complex space ℓ_2 is a Hilbert space, the scalar product of the elements $x = \{\xi_1 \dots \xi_n \dots\}$ and $y = \{\eta_1 \dots \eta_n \dots\}$ is defined by

$$\langle x, y \rangle = \sum_1 \xi_1 \bar{\eta}_1$$

Example 2: The complex space $L_2(a,b)$ becomes a Hilbert space, if the scalar product of the elements $x = x(t)$ and $y = y(t)$ is defined by $\langle x, y \rangle = \int_a^b x(t) \bar{y}(t) dt$

AI-7 Cartesian Product (of sets): If X_1 ($i = 1 \dots n$) is an element of the non empty set and $X = X_1 \times X_2 \times \dots \times X_n$, then the set defined by $X = \{ (x_1 \dots x_n) : x_i \in X_i \}$ is called the Cartesian product (of sets).

AI-8 Operator: (function) Let there be given two arbitrary sets X and Y and a rule by which there is associated with every $x \in X$ a uniquely determined element $y \in Y$, then we say that there is an operator (or abstract function) represents either by $y = A(x)$ or $y = Ax$ defined on the set X and whose range lies in the set Y , and often denote by $A : X \rightarrow Y$. An operator is often called a mapping, transformation, or function, since the operator can be thought of as carrying a point $x \in X$ into the point $y \in Y$.

AI-9 Range and Domain of Operator: Let $A : X \rightarrow Y$, the range of the operator A , denoted by R_A , is defined as $R_A = \{ y : y \in Y, y = Ax, x \in X \}$. If A is defined only on a subspace or subset of X , then the notation D_A is used to denote this set. D_A is called the domain of A . A particular subspace of the domain, called the null space of a linear operator, is the most interesting and important one, and is defined as follows:

AI-10 The Null Space of Linear Operator $A : X \rightarrow Y$ is the set

$$N_A = \{ x : x \in X, Ax = 0 \}$$

since $0 \in N_A$ it is clear that N_A is not void.

AI-11 Linear Operator: given two linear spaces X and Y , and let $y = Ax$ be an operator defined on X with a range lying in Y , then the operator A is called linear if

$$(1) A(x_1 + x_2) = Ax_1 + Ax_2$$

$$(2) A(\lambda x) = \lambda Ax$$

where λ is a scalar.

AI-12 Bounded Operator: An operator A is called bounded if there exists a constant M such that $\|Ax\| \leq M \|x\|$ for all $x \in X$.

AI-13 Functional: If the range of an operator consists of the underlying scalar field then the operator is called a functional. The space of linear functional defined on X is called the conjugate space of X and is designated by X^* . Hence $f \in X^*$, $f : X \rightarrow Y$ where Y is the spaces of scalar with dimension one. For example: let $X = L_2(0,1)$ then $f(x) = \int_0^1 x(t)dt$ is a linear functional which assigns to every $x(t) \in L_2(0,1)$ a real number. If we define $\langle x, y \rangle = \int_0^1 x(t)y(t)dt$ it is evident that $f(x) = \langle x, 1 \rangle$.

AI-14 Representation of Linear Functionals on Hilbert Space: Every linear functional $f(x)$ defined on a Hilbert space has the form $f(x) = \langle x, u \rangle$ with $\|f\| = \|u\|$ where u is uniquely determined by $f(x)$. If H is real space and if f is identical with u then we have

$$f(x) = \langle f, x \rangle$$

where for convenience we write $f(x) = \langle f, x \rangle$ rather than

$$f(x) = \langle x, f \rangle$$

AI-15 Adjoint Operator: Consider the linear operator $y = Ax$ which maps the Banach space E into E_1 . Let $g(y)$ be a linear functional defined on E_1 , i.e., $g(y) \in E_1^*$. Apply

the functional g to the element $y = Ax$ $g(Ax)$ is a linear functional defined on E ; denote it by $f(x)$. Thus $f(x) \in E^*$. We have assigned to each functional $g \in E_1^*$ a functional $f \in E^*$, i.e., we have obtained an operator which maps E_1^* into E^* . This operator is called the adjoint operator of A and is denoted by A^* . If we use the notation $\langle f, x \rangle$ for the functional $f(x)$, we obtained $\langle g, Ax \rangle = \langle A^*g, x \rangle$. If E and E_1 are both Hilbert spaces, then $E^* = E$, $E_1^* = E_1$, we have

$$\langle g, Ax \rangle_{E_1} = \langle A^*g, x \rangle_E$$

AI-16 Self Adjoint Operator:

A bounded linear operator A is called a bounded self adjoint (or Hermitian) operator if $A^* = A$.

AI-17 Positive Operator:

A self adjoint operator A different from the zero operator is called positive, $A > 0$, if

$$\langle x, Ax \rangle \geq 0$$

for all $x \in X$. It can be shown that if A is a positive operator there exists an operator K such that $A = K^*K$.

APPENDIX AII

THE DYADIC REPRESENTATION

The dyadic notation affords considerable insight into the representation of linear operators.

The brackets " $\langle$ " and " $\rangle$ " in $\langle f, u \rangle$ has the following meaning: we attach the symbol $\rangle$ to vectors and $\langle$ to the functionals thus $\langle f, u \rangle$ reads the functional $\langle f$ acting on the vector $u \rangle$.

Example 1. In the space $E_n : X = E_n$, the functional $f(x)$ on E_n can be represented by

$$f(x) = \langle f, x \rangle = \sum_{i=1}^n f_i x_i \quad (\text{AII-1})$$

where the x_i are the components of the vector $x \in E_n$ and the f_i arbitrary numbers. The f_i themselves can be interpreted as components of n -dimensional vector $f \in E_n^*$.

Example 2. In the space $L_p(T) : X = L_p(T)$, the functional $f(x)$ on $L_p(T)$ are of the form

$$f(x) = \langle x, u \rangle = \int_T u(s) x(s) ds \quad (\text{AII-2})$$

where

$$x(t) \in L_p(T)$$

$$u(t) \in L_q(T) \text{ is uniquely determined by } f \in L_p^*(T)$$

$$\frac{1}{p} + \frac{1}{q} = 1 \text{ for } 1 \leq p \leq \infty$$

$$\| f(x) \| = \| u \| = \left(\int_T |u(t)|^q dt \right)^{\frac{1}{q}}$$

If X and Y are linear spaces we shall interpret the expression

$$A = e \rangle \langle f \quad (\text{AII-3})$$

where $\langle f \in X^*$

$$e \rangle \in Y$$

as a linear operator $A \in L(X, Y)$ with range

$R_A = L(e \rangle) \subset Y$ and defined for any $x \in X$ by

$$A(x) = e \rangle \langle f, x \rangle \quad (\text{AII-4})$$

The operator A is linear and bounded if $\langle f$ is a bounded functional.

The representation of the operator A as given in AII-4 is called a one dimensional dyad. To generalize this result, consider the set $E = \{e_1 \dots e_n\} \subset Y$ and the set $F = \{\langle f_1 \dots \langle f_n\} \subset X^*$ and X is a normed linear space, then A is written as

$$A = \sum_{i=1}^n e_i \rangle \langle f_i$$

and defined by

$$A(x) = \sum_{i=1}^n e_i \rangle \langle f_i, x \rangle \quad \text{for } x \in X \text{ is called}$$

an n th order dyad. It is evident that the operator A is linear and bounded if the functional $\langle f_1 \dots \langle f_n$ are bounded, and the range of A is the linear manifold spanned by $e_1 \rangle \dots e_n \rangle$ and denoted by

$$R_A = L(e_1 \rangle \dots e_n \rangle)$$

If the set $\{e_i\}_1^n$ and $\{f_i\}_1^n$ are both linearly independent then the operator of the form

$A = \sum_{i=1}^n e_i \rangle \langle f_i$ will be called P-normal.

APPENDIX AIII

THE OPERATOR REPRESENTATION OF LINEAR DYNAMIC SYSTEM

AIII-1 Continuous Linear System

Consider a linear continuous dynamic system described by a time varying system differential equation

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \quad x(t_0) = x_0 \quad (\text{AIII-1.1})$$

where $t \in \mathcal{T} = [t_0, t_f]$, $A(t)$ and $B(t)$ are $n \times n$, and $n \times m$ real matrices defined on $\mathcal{T}$ and $u(t)$ and $x(t)$ are vectors with m and n components respectively

$$u(t) = [u_1(t) \dots u_m(t)]$$

$$x(t) = [x_1(t) \dots x_n(t)] \quad t \in \mathcal{T}$$

Let us assume that the components of $B(t)$ are bounded and (Lebesgue) measurable on $\mathcal{T}$, and that the norm of $A(t)$, considered as an operator on R^n , is (Lebesgue) integrable on $\mathcal{T}$, then, for each integrable input u there corresponds a unique absolutely continuous x satisfying AIII-1.1 almost everywhere on $\mathcal{T}$. It is well known (20) that under the above stated conditions on $A(t)$ and $B(t)$ the system AIII-1.1 has a unique solution given by

$$x(t) = \Phi(t, t_0)x_0 + \Phi(t, t_0) \int_{t_0}^t \Phi(t_0, s)B(s)u(s)ds$$

for $t, s \in \mathcal{T}$ (AIII-1.2)

where $\Phi(t, t_0)$ is called the transition matrix which is satisfied by

$$\dot{\Phi}(t, t_0) = A(t) \Phi(t, t_0)$$

$$\Phi(t_0, t_0) = I$$

where I denotes the identity matrix.

Equation AIII-1.2 yields an immediate characterization of the continuous system as a linear transformation on a Banach space.

Let us choose m number p_i ($1 \leq p_i \leq \infty$), denoted by p the ordered set $(p_1 \dots p_m)$, and let U_p denote the cartesian product of the corresponding spaces

$$U_p = L_{p_1} \times L_{p_2} \times \dots \times L_{p_m}$$

where L_{p_1} is the Banach space of functions with integrable p_1 -th power. Let us assume that each $u(t)$ is a member of the Banach space L_{p_1} , then the total input $u(t) = [u_1(t) \dots u_m(t)]$ is a member of the tuplet space $U_p = L_{p_1} \times L_{p_2} \times \dots \times L_{p_m}$. If we treat $v = (x_0, u)$ as an element of the input space V_p , defined by $V_p = R^n \times U_p$, then for each $v = (x_0, u) \in V_p$ there is a unique absolutely continuous x satisfying equation AIII-1.1. Let us define an operator T on V_p that maps v into x , then we have

$$x(t) = (Tv)(t) = \Phi(t, t_0)x_0 + \Phi(t, t_0) \int_{t_0}^t \Phi(t_0, s)B(s)u(s)ds$$

Since one is often only interested on certain specified time $x(t)$ of the system output, the system acts on its input space V_p to produce a n -tuple of real numbers. From this point of

view we may define an operator T^t such that

$$x(t) = T^t v = (Tv)(t)$$

where t is fixed in τ , and $x(t)$ is the t -th sample of the output function x . The T determines in this way a family of mapping $\{T^t : t \in \tau\}$, each of which is defined on V_p having values in R^n .

It is shown in Appendix AIV and AV that T and T^t are bounded linear operators. Therefore, in summary, every continuous linear system can be represented as a family $\{T^t : t \in \tau\}$ of bounded linear operators, T^t , each of which is defined on V_p having values in R^n . These mapping are indexed by the base $t \in \tau$ on which the input function is defined. T^t acts on the total system input to produce the t -th sample of the output.

The operator T^t has a dyadic representation which can be formulated in the following way:

Let $Z_1(t) = (Z_{11}(t) \dots Z_{1m}(t))$ denote the i th row of $Z(t)$ defined by $Z(t) = \Phi(t_0, t)B(t)$ which is a bounded $n \times m$ matrix and let

$$[Z_1, u](s) = \sum_{j=1}^m Z_{1j}(s)u_j(s)$$

then we have

$$\int_{t_0}^t \Phi(t_0, s)B(s)u(s)ds = \int_{t_0}^t \begin{bmatrix} [Z_1, u](s) \\ [Z_2, u](s) \\ \vdots \\ [Z_n, u](s) \end{bmatrix} ds \quad (\text{AIII-1.3})$$

Moreover, the elements of $Z(t)$ are bounded and measurable since B has these properties and $\bar{\Phi}(t, t_0)$ is continuous.

In particular $Z_{1j} \in L_{q_j}$ where $u_j \in L_{p_j}$, $1 \leq p_j \leq \infty$ and $\frac{1}{p_j} + \frac{1}{q_j} = 1$ and hence the mapping which sends u_j into the member

$\int_{t_0}^t Z_{1j}(s)u_j(s) ds$ can be considered as a bounded linear functional on L_{p_j} . Let it be designated by r_{1j}^t and defined

as

$$r_{1j}^t(u_j) \triangleq \int_{t_0}^t Z_{1j}(s)u_j(s)ds \quad (\text{AIII-1.4})$$

and the functional $r_1^t \in U^*$ can be defined in the natural way

$$r_1^t(u) = \sum_{j=1}^m r_{1j}^t(u_j) \quad (\text{AIII-1.5})$$

or in the dyadic notation

$$r_1^t(u) \triangleq \langle r_1, u \rangle$$

then the force response in equation AIII-1.1 can be written

as

$$\bar{\Phi}(t, t_0) \int_{t_0}^t \bar{\Phi}(t_0, s) B(s) u(s) ds = \sum_{i=1}^n \phi_i(t, t_0) \langle r_i, u \rangle = F^t u \quad (\text{AIII-1.6})$$

where the family of operators F^t is defined by

$$F^t \triangleq \sum_{i=1}^n \phi_i(t, t_0) \langle r_i \quad t \in \tau \quad (\text{AIII-1.7})$$

that maps U_p into R^n .

Let us consider the initial condition response of the system this term is given by $\bar{\Phi}(t, t_0)x_0$. Let R^n denote the space of

real n -tuplets, the e_1^* defined by

$$e_1^*(x_0) = x_{10} \quad (\text{AIII-1.8})$$

defines a bounded linear functional on R^n , then

$$\Phi(t, t_0)x_0 = \sum_{i=1}^n \phi_i(t, t_0) \langle e_i^*, x_0 \rangle = J^t x_0$$

where the family of operators J^t is defined by

$$J^t \triangleq \sum_{i=1}^n \phi_i(t, t_0) \langle e_i^* \rangle, \quad t \in \tau \quad (\text{AIII-1.9})$$

that maps R^n into R^n . Thus in the continuous case, the output $x(t)$ and hence $T^t v$ can be written

$$\begin{aligned} x(t) = T^t v &= J^t x_0 + F^t u = \sum_{i=1}^n \phi_i(t, t_0) \langle e_i^*, x_0 \rangle \\ &\quad + \sum_{i=1}^n \phi_i(t, t_0) \langle f_i, u \rangle \\ &= \sum_{i=1}^n \phi_i(t, t_0) \langle g_i, v \rangle, \quad t \in \tau \quad (\text{AIII-1.10}) \end{aligned}$$

where $T^t = J^t \oplus F^t$

$$\langle e^* \in (R^n)^*$$

and $\langle g_i = \langle (e_i^*, f_i) \rangle$ defines a linear bounded functional on V_p .

Hence we may write

$$T^t = \sum_{i=1}^n \phi_i(t, t_0) \langle g_i \quad (\text{AIII-1.11})$$

In conclusion we have the following result:

If F^t is the parametrized family of bounded linear operators that represents the forced response of the system AIII-1.1, and if J^t is the bounded linear operator representing its

free response, then F^t and J^t can be written in the form of equation AIII-1.7 and AIII-1.9 respectively. Here $\phi_1(t, t_0)$ are the columns of the state transition matrix $\Phi(t, t_0)$, $\langle e_1^* \rangle$ is the linear functional on R^n defined by AIII-1.8 and f_1^t is given by AIII-1.5 for $t = \tau$. Finally $\langle g_1 = \langle (e_1^*, f_1) \rangle$ defines a bounded linear functional on V_p . We see that

$$\tau^t = J^t (+ F^t = \sum_{i=1}^n \phi_i(t, t_0) \rangle \langle g_i$$

gives a decomposition of the total system response in terms of the basis $\{\phi_1 \dots \phi_n\}$ of R^n and n -bounded linear functional $\langle g_i$ on the input space V_p .

AIII-2 Discrete Linear System

Consider a linear discrete dynamic system described by the difference equation

$$\Delta_k x(t_k) = A(t_k)x(t_k) + B(t_k)u(t_k) \quad ; \quad x(t_0) = x_0 \quad (\text{AIII-2.1})$$

where $t_k \in \mathcal{G} = [t_0, t_1 \dots t_k \dots t_f]$

The matrices $A(t_k)$ and $B(t_k)$ are $n \times n$ and $n \times m$ real matrices defined on $\mathcal{G}$, $u(t_k)$ and $x(t_k)$ are vectors with m and n components respectively, each of which is defined on $\mathcal{G}$. The difference operator Δ_k is defined by

$$\Delta_k x(t_k) = h_k^{-1} [x(t_{k+1}) - x(t_k)] \quad \text{for } t = t_k \quad (\text{AIII-2.2})$$

$$h_k = t_{k+1} - t_k$$

The existence and uniqueness of the solution x to AIII-2.1

is well known and the solution has the form

$$x(t_k) = \Phi(t_k, t_0)x_0 + \Phi(t_k, t_0) \sum_{\alpha=0}^{k-1} h_{\alpha} \Phi(t_0, t_{\alpha+1})B(t_{\alpha})u(t_{\alpha}) \quad (\text{AIII-2.3})$$

for $t_k \in \delta$. The $n \times n$ matrix sequence $\{\Phi(t_k, t_0)\}$ is called the transition matrix sequence for AIII-2.1 and satisfies the associated matrix differential equation

$$\Delta_k \Phi(t_k, t_0) = A(t_k) \Phi(t_k, t_0)$$

$$\Phi(t_0, t_0) = I \quad (\text{AIII-2.4})$$

for all $t_k \in \delta$. Following the same manner as for the case of the continuous system, every discrete linear dynamic system can be represented as a bounded linear transformation T from one Banach space V into another.

For discrete systems, we let U be the m -fold cartesian product of $\ell(\delta)$ where $\ell(\delta)$ is the vector space of all real-valued function defined on

$$U = \ell(\delta) \times \ell(\delta) \times \dots \times \ell(\delta) \quad (m \text{ copies})$$

Let $V = \mathbb{R}^n \times U$ be the vector space of input $v = (x_0, u)$ to the discrete system. Define the transformation $T : V \rightarrow U$ by requiring that image of $v(t_k)$ be this $x(t_k)$:

$$(Tv)(t_k) = x(t_k)$$

where $x(t_k)$ is given by AIII-2.3.

Here T is well defined, linear and bounded (since V is finite dimensional) relative to any pair of norms on V and U . Using the same argument as for the case of continuous systems, one is often interested only in the certain specified samples $x(t_k)$ of the system output. The system acts on its input

space V to produce an n -tuple of real numbers. One may consider the system as a linear operator into R^n and the mapping T^{t_k} defined by

$$T^{t_k} v = (Tv)(t_k) \quad \text{for fixed } t_k \in \delta \quad (\text{AIII-2.5})$$

as a family of bounded linear operator each of which is defined on V having values in R^n . These mapping are indexed by the base set $t \in \delta$ on which the input function are defined; T^{t_k} acts on the total system input to produce the t_k -th sample of the output.

By the same argument the family operator T^{t_k} in the discrete system has the similar dyadic representation as that T^t in the continuous system. Let $Z_{1j}(t_\ell)$ be the (i,j) components of $Z(t_\ell)$ defined by

$$Z(t_\ell) = \Phi(t_0, t_{\ell+1})B(t_\ell)$$

The mapping $f_1^{t_k} \in U^* \quad i = 1, \dots, n$

where $U^* = \ell_{p_1}^*(\delta) \times \ell_{p_2}^*(\delta) \times \dots \times \ell_{p_m}^*(\delta)$

sends the element $u_j(t_\ell)$ into the number

$$f_j^{t_k}(u) = \sum_{\ell=0}^{k-1} \sum_{j=1}^m h_\ell z_{1j}(t_\ell) u_j(t_\ell) = \langle f_j, u \rangle \quad (\text{AIII-2.6})$$

which is linear and bounded operator defined on U (since U is finite-dimensional). By use of the dyadic notation the force response of the system AIII-2.1 can be represented by

$$\begin{aligned} \sum_{i=1}^n \phi_i(t_k, t_0) \sum_{\ell=0}^{k-1} \sum_{j=1}^m h_\ell z_{1j}(t_\ell) u_j(t_\ell) = \\ \sum_{i=1}^n \phi_i(t_k, t_0) \langle f_i, u \rangle = F^{t_k} u \end{aligned} \quad (\text{AIII-2.7})$$

where the n -dimensional dyadic operator F^{tk} is defined by

$$F^{tk} \triangleq \sum_{i=1}^n \phi_1(t_k, t_0) > < f_i \quad (\text{AIII-2.8})$$

Similarly, the initial condition response of the system has the form

$$\phi(t_k, t_0)x_0 = \sum_{i=1}^n \phi_1(t_k, t_0) > < e_i^*, x_0 > = J^{tk}x_0 \quad (\text{AIII-2.9})$$

$$\text{where } J^{tk} \triangleq \sum_{i=1}^n \phi_1(t_k, t_0) > < e_i^* \quad (\text{AIII-2.10})$$

The total response of the system has the operator representation

$$\begin{aligned} x(t_k) &= T^{tk}v = J^{tk}x_0 + F^{tk}u = \sum_{i=1}^n \phi_1(t_k, t_0) > \\ &\quad [< e_i^*, x_0 > + < f_i, u >] \\ &= \sum_{i=1}^n \phi_1(t_k, t_0) > < g_i, v > \quad \text{for } t_k \in \delta \quad (\text{AIII-2.11}) \end{aligned}$$

$$\text{Hence we may write } T^{tk} = \sum_{i=1}^n \phi_1(t_k, t_0) > < g_i$$

where $< e_i^* \in (R^n)^*$

$$< f_i \in U^*$$

and $< g_i = < (e_i^*, f_i)$ defines a linear bounded functional on V .

Suppose that the set δ is ordered so that $t_{j-1} < t_j$ and defines "the greatest t_k function" $[t]$ by

$$[t] = \max_{t_k \in \delta} \left\{ k : t_k \leq t \right\} \quad (\text{AIII-2.12})$$

then $[t] = k$ if and only if $t \in [t_k, t_{k+1}]$.

The equation AIII-2.3 can be written as

$$x(t) = \Phi(t, t_0)x_0 + \Phi(t, t_0) \sum_{j=0}^{[t]-1} h_j \Phi(t_0, t_{j+1}) \\ B(t_j)u(t_j) \quad , \quad t \in \delta$$

and equation AIII-2.11 becomes

$$x(t) = \sum_{i=1}^n \phi_i(t, t_0) > \langle g_i, v \rangle \quad t \in \delta$$

which has the same form as that in the continuous case.

APPENDIX AIV

PROOF THAT $(Tv)(t)$ DEFINED BY

$$(Tv)(t) = \Phi(t, t_0)x_0 + \Phi(t, t_0) \int_{t_0}^t \Phi(t_0, s)B(s)u(s)ds$$

(AIV-1)

IS A LINEAR AND BOUNDED OPERATOR

Linearity of T:

It is obvious that T is a linear operator because

$$\begin{aligned} T(\alpha v + \beta w) &= \Phi(t, t_0)\alpha x_0 + \Phi(t, t_0) \int_{t_0}^t \Phi(t_0, s)B(s)\alpha u(s) ds \\ &\quad + \Phi(t, t_0)\beta x_0 + \Phi(t, t_0) \int_{t_0}^t \Phi(t_0, s)B(s)\beta z(s) ds \\ &= \alpha \left[\Phi(t, t_0)x_0 + \Phi(t, t_0) \int_{t_0}^t \Phi(t_0, s)B(s)u(s) ds \right] \\ &\quad + \beta \left[\Phi(t, t_0)x_0 + \Phi(t, t_0) \int_{t_0}^t \Phi(t_0, s)B(s)z(s) ds \right] \\ &= \alpha T(v) + \beta T(w) \end{aligned}$$

where $v = (x_0, u)$ and $w = (x_0, z)$.

Boundedness of T:

Consider 1-th component of Tv that is T_1v and treat T_1 as a mapping into $C(t)$. From equation AIV-1

$$T_1v = \phi_1(t, t_0)x_{1_0} + \phi_1(t, t_0) \int_{t_0}^t \sum_{j=1}^m w_{1j}(s)u_j(s) ds$$

where w_{1j} is the $(1, j)$ element of the matrix $\Phi(t_0, s)B(s)$

$$\begin{aligned}
|T_1 v| &= |\phi_1(t, t_0)x_{10} + \phi_1(t, t_0) \int_{t_0}^t \sum_{j=1}^m w_{1j}(s)u_j(s) ds| \\
&\leq |\phi_1(t, t_0)| |x_{10}| + |\phi_1(t, t_0)| \sum_{j=1}^m \int_{t_0}^t w_{1j}(s)u_j(s) ds
\end{aligned}$$

By the application of Hölder's inequality we have

$$|T_1 v| \leq |\phi_1(t, t_0)| |x_{10}| + |\phi_1(t, t_0)| \sum_{j=1}^m \|w_{1j}\|_{q_j} \|u_j\|_{p_j} \quad (\text{AIV-2})$$

where $\|u_j\|_{p_j}$ is the norm of $u_j \in L_{p_j}$, $\|w_{1j}\|_{q_j}$ is the norm of $w_{1j} \in L_{q_j}$, and $\frac{1}{p_j} + \frac{1}{q_j} = 1$. Equation (AIV-2) may be

written as

$$\begin{aligned}
|T_1 v| &\leq |\phi_1(t, t_0)| \left\{ |x_{01}| + \max_{1,j} \|w_{1j}\|_{q_j} \cdot \sum_{j=1}^m \|u_j\|_{p_j} \right\} \\
&\leq \max_{1,j} \left\{ 1, \max_{1,j} \|w_{1j}\|_{q_j} \right\} \cdot |\phi_1(t, t_0)| \cdot \left\{ |x_{01}| + \sum_{j=1}^m \|u_j\|_{p_j} \right\}
\end{aligned}$$

so that

$$\|T_1 v\| = \sup_{t \in \tau} |T_1 v(t)| \leq M \left\{ |x_{01}| + \sum_{j=1}^m \|u_j\| \right\}$$

where

$$M = \max_{1,j} \left\{ 1, \max_{1,j} \|w_{1j}\|_{q_j} \right\} \cdot \sup_{t \in \tau} |\phi_1(t, t_0)|$$

Hence T_1 is bounded. The boundedness of T follows from the fact that a mapping into a product of Banach space is bounded if and only if its coordinate maps are bounded, i.e.,

$T: R^n_x U \rightarrow C^n$ is bounded if and only if the mappings $T_1: R^n_x U \rightarrow C^n$ defined by $T_1 v = 1$ -th component of Tv are bounded.

APPENDIX AV

PROOF THAT F^t DEFINED BY

$$F^t u = \int_{t_0}^t \Phi(t,s) B(s) u(s) ds$$

IS BOUNDED FOR FIXED t IN

THE INTERVAL τ , WHERE

$$F^t : L_2^m(\tau) \rightarrow R^n$$

Let the norm on R^n for $\|v\| = \sqrt{v_1^2 + \dots + v_n^2}$ be denoted by $\|v\|_{R^n}$ and for $\|v\| = \max_i |v_i|$ be denoted by

$\|v\|_{R^n_{\max}}$. Let the norm on R^n be defined by $\|v\| = \sqrt{v_1^2 + \dots + v_n^2}$,

and consider

$$\begin{aligned} \|F^t u\|_{R^n_{\max}} &= \left\| \int_{t_0}^t \Phi(t,s) B(s) u(s) ds \right\|_{R^n_{\max}} \\ &= \max_i \left| i\text{-th component of } \int_{t_0}^t \Phi(t,s) B(s) u(s) ds \right| \\ &= \max_i \left| \int_{t_0}^t \sum_k \sum_j \Phi_{1,j}(t,s) B_{j,k}(s) u_k(s) ds \right| \\ &\leq \max_i \left\{ \int_{t_0}^t \sum_k \left| \sum_j \Phi_{1,j}(t,s) B_{j,k}(s) u_k(s) \right| ds \right\} \end{aligned}$$

By the application of Schwarz inequality we have

$$\|F^t u\|_{R^n_{\max}} \leq \max_i \left\{ \left[\int_{t_0}^t \sum_k \left| \sum_j \Phi_{1,j}(t,s) B_{j,k}(s) \right|^2 ds \right]^{1/2} \left[\int_{t_0}^t \sum_k |u_k(s)|^2 ds \right]^{1/2} \right\}$$

$$= \max_1 \left\{ \left[\int_{t_0}^t \sum_k \left| \sum_j \Phi_{1,j}(t,s) B_{j,k}(s) \right|^2 ds \right]^{1/2} \right\} \\ \left[\int_{t_0}^t \sum_k |u_k(s)|^2 ds \right]^{1/2} \quad (\text{AV-1})$$

Since $\Phi(t,s)$ is a continuous function in the finite interval $[t_0, t]$, $B(t)$ is bounded in this interval and t is fixed, let this fixed number

$$\max_1 \left\{ \left[\int_{t_0}^t \sum_k \left| \sum_j \Phi_{1,j}(t,s) B_{j,k}(s) \right|^2 ds \right]^{1/2} \right\}$$

be denoted by M , then equation (AV-1) may be written as

$$\|F^t u\|_{R^n \max} \leq M \left[\int_{t_0}^t \sum_k |u_k(s)|^2 ds \right]^{1/2} \leq M \left[\int_{\tau} \sum_k |u_k(s)|^2 ds \right]^{1/2} \quad (\text{AV-2})$$

Since for $v \in R^n$ we have

$$\begin{aligned} \|v\|_{R^n} &= \sqrt{v_1^2 + \dots + v_n^2} \\ &\leq \sqrt{\max_1 |v_1|^2 + \dots + \max_1 |v_1|^2} \\ &= \sqrt{n \max_1 |v_1|^2} \\ &= \sqrt{n} \max_1 |v_1| \\ &= \sqrt{n} \|v\|_{R^n \max} \end{aligned} \quad (\text{AV-3})$$

From equations (AV-2) and (AV-3) we have

$$\frac{1}{\sqrt{n}} \|F^t u\|_{R^n \max} \leq \|F^t u\|_{R^n \max} \leq M \sqrt{\int_{\tau} \sum_k |u_k(s)|^2 ds} = M \|u\|_{L_2(\tau)}^{m(\tau)}$$

$$\text{or } \| F^t u \|_{R^n} \leq C \| u \|_{L_2^m(\tau)}$$

where $C = M\sqrt{n}$, and $\| u \|_{L_2^m(\tau)}$ is the norm on $L_2^m(\tau)$ space.

Remark: The boundedness of T^t , defined by

$$T^t v = \Phi(t, t_0)x_0 + \int_{t_0}^t \Phi(t, s)B(s)u(s) ds \text{ with fixed } x_0,$$

is immediate from the fact that $\Phi(t, t_0)x_0$ is a continuous function in the finite interval $[t_0, t]$, and every continuous function in a finite interval is bounded.

APPENDIX AVI

A PROOF OF

$$F^*x = B^*(t) \Phi^*(t_0, t) \int_t^{t_f} \Phi^*(s, t_0) x(s) ds, \quad t \in \mathcal{T}$$

From the definition of adjoint operator we have

$$\langle x, Fy \rangle = \langle F^*x, y \rangle \quad (\text{AVI-1})$$

In real space $L_2(t_0, t_f)$ the inner product of two elements u and v is defined by

$$\langle u, v \rangle = \int_{t_0}^{t_f} [u, v] ds$$

thus

$$\begin{aligned} \langle x, Fy \rangle &= \int_{t_0}^{t_f} [x(s_1), \Phi(s_1, t_0) \int_{t_0}^{s_1} \Phi(t_0, s_2) B(s_2) y(s_2) ds_2] ds_1 \\ &= \int_{t_0}^{t_f} \int_{t_0}^{s_1} y^*(s_2) B^*(s_2) \Phi^*(t_0, s_2) \Phi^*(s_1, t_0) x(s_1) ds_2 ds_1 \end{aligned} \quad (\text{AVI-2})$$

where

$$(Fy)(t) = \Phi(t, t_0) \int_{t_0}^t \Phi(t_0, s) B(s) u(s) ds$$

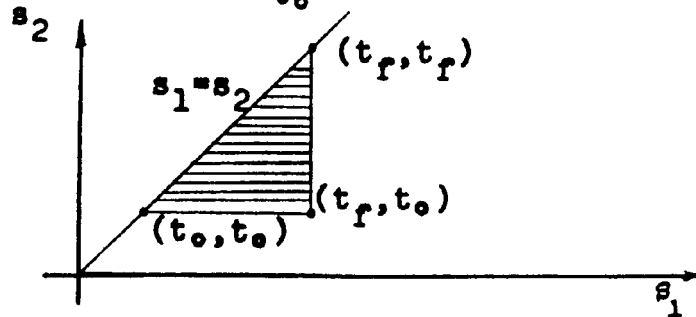


Fig. AVI-1.--Domain of integration

Now equation (VI-2) can be written as (see Figure AVI-1):

$$\langle x, Fy \rangle = \int_{t_0}^{t_f} \int_{s_2}^{t_f} y^*(s_2) B^*(s_2) \bar{\Phi}^*(t_0, s_2) \bar{\Phi}^*(s_1, t_0) x(s_1) ds_1 ds_2$$

or

$$\langle x, Fy \rangle = \left\langle \int_{s_2}^{t_f} B^*(s_2) \bar{\Phi}^*(t_0, s_2) \bar{\Phi}^*(s_1, t_0) x(s_1) ds_1, y(s_2) \right\rangle$$

(AVI-3)

From (AVI-1) and (AVI-3) we have

$$(F^*x)(s_2) = \int_{s_2}^{t_f} B^*(s_2) \bar{\Phi}^*(t_0, s_2) \bar{\Phi}^*(s_1, t_0) x(s_1) ds_1$$

or

$$(F^*x)(t) = B^*(t) \bar{\Phi}^*(t_0, t) \int_t^{t_f} \bar{\Phi}^*(s, t_0) x(s) ds$$

APPENDIX AVII

A PROOF OF

$$(F^{t_f})^* x(t_f) = B^*(t) \Phi^*(t_0, t) \Phi^*(t_f, t_0) x(t_f)$$

From the definition of adjoint operator we have

$$\langle x(t_f), F^{t_f} y \rangle = \langle (F^{t_f})^* x(t_f), y \rangle. \quad (\text{AVII-1})$$

In real space l_2 the inner product of two elements u and v is defined by

$$\langle u, v \rangle = \sum_{i=1}^n u_i v_i \quad (\text{AVII-2})$$

and in L_2 the inner product of two elements u and v is defined by

$$\langle u, v \rangle = \int_{t_0}^{t_f} [u(s), v(s)] ds \quad (\text{AVII-3})$$

also

$$F^{t_f} y = \sum_1 \phi_1 \langle \phi_1 \rangle = \Phi(t_f, t_0) \int_{t_0}^{t_f} \Phi(t_0, s) B(s) y(s) ds. \quad (\text{AVII-4})$$

Substituting (AVII-4) in (AVII-1) we have

$$\begin{aligned} \langle x(t_f), F^{t_f} y \rangle &= \langle x(t_f), \Phi(t_f, t_0) \int_{t_0}^{t_f} \Phi(t_0, s) B(s) y(s) ds \rangle \\ &= \int_{t_0}^{t_f} y^*(s) B^*(s) \Phi^*(t_0, s) \Phi^*(t_f, t_0) x(t_f) ds \\ &= \langle B^*(s) \Phi^*(t_0, s) \Phi^*(t_f, t_0) x(t_f), y(s) \rangle \end{aligned}$$

Hence

$$(F^{t_f})^* x(t_f) = B^*(t) \Phi^*(t_0, t) \Phi^*(t_f, t_0) x(t_f)$$

APPENDIX AVIII

A PROOF OF $\mathcal{F}^{-1} = [\langle f_1^+, f_j^+ \rangle]$

From (2.2.1-6) we have

$$\mathcal{F} = [\langle f_1, f_j \rangle]$$

$$= \begin{bmatrix} \langle f_1 \\ \vdots \\ \langle f_n \end{bmatrix} [f_1 \rangle \dots f_n \rangle]$$

where $\langle f_1$ is treated as a row vector and $f_j \rangle$ as a column vector. Using the rule for the inverse of the product of two matrices we have formally

$$\mathcal{F}^{-1} = [f_1 \rangle \dots f_n \rangle]^{-1} \begin{bmatrix} \langle f_1 \\ \vdots \\ \langle f_n \end{bmatrix}^{-1}$$

By definition of $\langle f_1^+, f_j^+ \rangle = \delta_{1j}$ in matrix form we have

$$\begin{bmatrix} \langle f_1^+ \\ \vdots \\ \langle f_n^+ \end{bmatrix} [f_1 \rangle \dots f_n \rangle] = [\langle f_1^+, f_j^+ \rangle] = I$$

which leads to the form of $\mathcal{F}^{-1}$

$$\mathcal{F}^{-1} = \begin{bmatrix} \langle f_1^+ \\ \vdots \\ \langle f_n^+ \end{bmatrix} [f_1^+ \rangle \dots f_n^+ \rangle] = [\langle f_1^+, f_j^+ \rangle]$$

Remark: By the same argument

$$\text{if } \mathcal{N} = [\langle h_1, h_j \rangle]$$

we have

$$\mathcal{N}^{-1} = [\langle h_1^+, h_j^+ \rangle]$$

APPENDIX AIX

$$\text{A PROOF OF } \|\delta x\|^2 \leq M^2 \int_{t_0}^{t_f} [\bar{x}^{\#} \delta A^{\#} \delta A \bar{x}] dt$$

We have

$$\delta x(t) = \int_{t_0}^t \Phi(t, \tau) \delta A(\tau) \bar{x}(\tau) d\tau$$

$$\text{Let } y(\tau) = \delta A(\tau) \bar{x}(\tau)$$

$$\|\delta x\|^2 = \int_{t_0}^{t_f} \sum_{i=1}^n (\delta x)_i^2 dt = \int_{t_0}^{t_f} \sum_{i=1}^n \left[\int_{t_0}^t \sum_{j=1}^n w_{ij}(t, s) y_j(s) ds \right]^2 ds$$

(AIX-1)

By use of Schwarz inequality

$$\sum_{i=1}^n a_i b_i \leq \sqrt{\sum_{i=1}^n a_i^2 \sum_{i=1}^n b_i^2}$$

$$\text{and } \left[\int_{t_0}^{t_f} f g dt \right] \leq \left(\int_{t_0}^{t_f} f^2 dt \int_{t_0}^{t_f} g^2 dt \right)^{1/2}$$

equation (AIX-1) reduces to

$$\begin{aligned} \|\delta x\|^2 &\leq \int_{t_0}^{t_f} \sum_{i=1}^n \left[\int_{t_0}^t \sqrt{\sum_{j=1}^n w_{ij}^2(t, s) \sum_{j=1}^n y_j^2(s)} ds \right]^2 dt \\ &\leq \int_{t_0}^{t_f} \sum_{i=1}^n \left[\int_{t_0}^t \sum_{j=1}^n w_{ij}^2(t, s) ds \int_{t_0}^t \sum_{j=1}^n y_j^2(s) ds \right] dt \\ &\leq \int_{t_0}^{t_f} \sum_{i=1}^n \left[\int_{t_0}^t \sum_{j=1}^n w_{ij}^2(t, s) ds \int_{t_0}^{t_f} \sum_{j=1}^n y_j^2(s) ds \right] dt \end{aligned}$$

$$= \int_{t_0}^{t_f} \int_{t_0}^t \sum_{i,j=1}^n w_{ij}^2(t,s) ds dt \int_{t_0}^{t_f} [\bar{x}^* \delta A^* \delta A \bar{x}] ds$$

Hence

$$E \|\delta x\|^2 \leq M^2 \int_{t_0}^{t_f} [\bar{x}^* E(\delta A^* \delta A) \bar{x}] dt = \int_{t_0}^{t_f} [\bar{x}, Q \bar{x}] dt$$

where

$$Q = M^2 E[\delta A^* \delta A] \quad , \quad M^2 = \int_{t_0}^{t_f} \int_{t_0}^t \sum_{i,j} w_{ij}^2(t,s) ds dt,$$

w_{ij} is the (i,j) element of the transition matrix $\Phi(t,\tau)$ and $\bar{x}$ is the mean value of x .

APPENDIX AX

THE METHOD OF STEEPEST DESCENT

In Chapter I, we introduced the concept of steepest descent method in the minimization of $z=f(x)$, $x=(x_1 \dots x_n)$ in Hilbert space. The sequence at $j+1$ step is given by

$$x^{(j+1)} = x^{(j)} - k_j \nabla f(x^{(j)})$$

To appreciate this fact it is useful to go into the geometry of the problem. The function $z=f(x)$ with $x=(x_1 \dots x_n)$ represents a surface in $n+1$ dimension and $f(x)=c$, where c is a constant, expresses a contour on a plane surface which is perpendicular to the z axis and cuts it at $z=c$. By varying c , we obtain a family of contours. Since the gradient at a point is perpendicular to any contour passing through the point, we seek the minimum (or maximum) by moving along the gradient from one contour surface to the next (see figure AX-1).

The method of choosing the coordinates of the new point is therefore to subtract a unit gradient from the coordinates of the previous point. To obtain the best convergence, we could have used scalar multiples of the gradient. This scalar factor k_j at j -th step may be determined by forming a Taylor series of $f(x^{(j+1)})$ about the point $x^{(j)}$. If we keep three terms in this series expansion then from

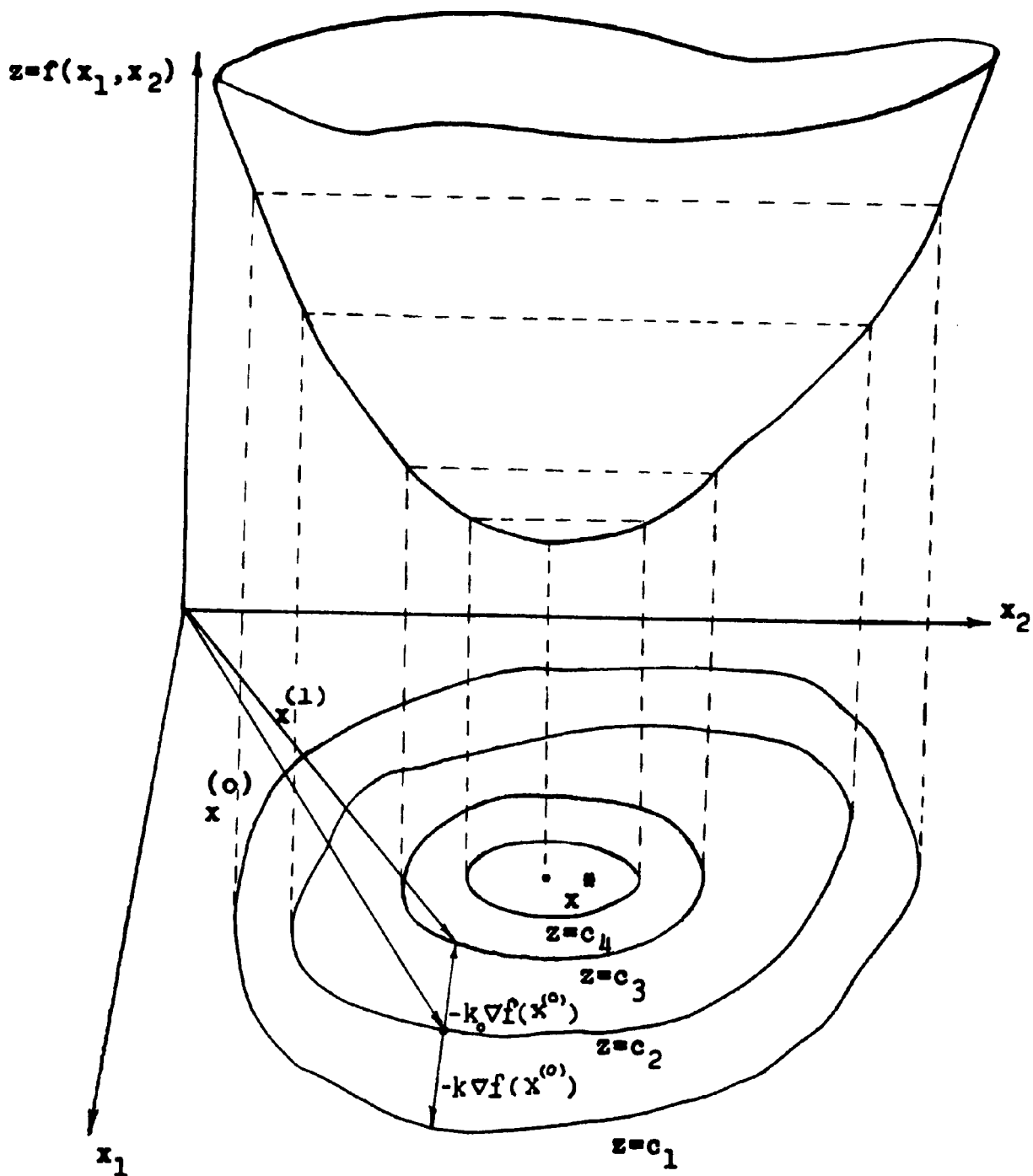


Fig. AX-1.--A geometric interpretation of the method of steepest descent in the minimization of $z=f(x_1, x_2)$.

the condition of minimum of $f(x^{(j+1)})$ it gives the value of k_j . Since evaluating the second derivative in the determination of k_j along the given coordinate can be time consuming, an approximation to the optimum constant is frequently used.

In this chapter two types of problems which are closely related to those given in this dissertation are solved by the method of steepest descent and are discussed. We mainly deal with:

(1) The solution to the operator equation $y=kx$ where k is bounded linear operator which is applicable to the solution of $(I + F^*F)u=v$ mentioned in Chapter II, and

(2) The solution to the minimization of $f(x)$ with the inequality constraint $g(x) \leq 0$, $x = (x_1 \dots x_n)$ which will be applied to the determination of parameters $\{f_i\}_1^n$ from

$J^0(\xi)$ subject to the constraint $\xi_1^2 + \dots + \xi_n^2 \leq c^2$.

AX-1. The Solution to the Operator Equation $y = Kx$

Let H denote a real Hilbert space and consider the operator equation $L(x) = Kx - y = 0$ (AX-1.1)

where x and y are elements of the Hilbert space H and K is a linear operator with domain H and range R_K . Let us assume that K is a symmetric positive definite operator, that is

$\langle x, Kx \rangle \gg 0$ for all $x \in H$. Given a $y \in R_K$ find $x \in H$ such that $y = Kx$. We introduce the equation functional

$$f(x) = \langle Kx, x \rangle - 2 \langle x, y \rangle \quad (AX-1.2)$$

The minimum point of $f(x)$ occurs at the point x^* if $\nabla f_{x^*} = 0$.

Directly from equation (AX-1.2) we see that

$$\nabla f_x = 2(Kx - y)$$

from this it is clear that the problem of solving equation (AX-1.1) reduces to the problem of seeking the minimum of the functional (AX-1.2).

Let z be an element of the space H and k a real parameter then we have

$$\begin{aligned} f(x+kz) &= \langle Kx+kKz, x+kz \rangle - 2 \langle x+kz, y \rangle \\ &= \langle Kx, x \rangle + k[\langle Kx, z \rangle + \langle Kz, x \rangle] + k^2 \langle Kz, z \rangle - 2 \langle x, y \rangle - 2k \langle z, y \rangle \\ &= f(x) + 2k \langle Kx - y, z \rangle + k^2 \langle Kz, z \rangle \quad (AX-1.3) \end{aligned}$$

We shall select z from equation (AX-1.3) in the following manner (see equation 1.2-4):

$$z^{(0)} = - \frac{\nabla f_{x^{(0)}}}{\|\nabla f_{x^{(0)}}\|} = - \frac{L(x^{(0)})}{\|L(x^{(0)})\|}$$

where $\|z^{(0)}\| = 1$ and $\nabla f_{x^{(0)}} = 2 \langle Ax^{(0)}, \phi \rangle = 2 L(x^{(0)})$

Since (AX-1.3) is a second order polynomial in k we can determine k from the condition of a minimum of $f(x^{(0)} + kz)$ that is the value of k_0 for which

$$\frac{d}{dk} f(x^{(0)} + kz^{(0)}) = 0$$

This result is shown to be

$$k_0 = \frac{-\langle Kx^{(0)} - y, z^{(0)} \rangle}{\langle Kz^{(0)}, z^{(0)} \rangle} = \frac{\langle L(x^{(0)}), \frac{L(x^{(0)})}{\|L(x^{(0)})\|} \rangle}{\langle K \frac{L(x^{(0)})}{\|x^{(0)}\|}, \frac{L(x^{(0)})}{\|x^{(0)}\|} \rangle}$$

Collecting these results yield the expression

$$k_0 z^{(0)} = - \frac{\langle L(x^{(0)}), L(x^{(0)}) \rangle}{\langle KL(x^{(0)}), L(x^{(0)}) \rangle} L(x^{(0)}) \quad (\text{AX-1.4})$$

Hence the element $x^{(1)}$ which is adopted as the next approximation is given by

$$x^{(1)} = x^{(0)} + k_0 z^{(0)} = x^{(0)} - \frac{\langle L(x^{(0)}), L(x^{(0)}) \rangle}{\langle KL(x^{(0)}), L(x^{(0)}) \rangle} L(x^{(0)}) \quad (\text{AX-1.5})$$

By repeating the above procedure, it is evident that at the n -th step the approximation $x^{(n)}$ is given by

$$x^{(n)} = x^{(n-1)} - \frac{\langle w^{(n-1)}, w^{(n-1)} \rangle}{\langle Kw^{(n-1)}, w^{(n-1)} \rangle} w^{(n-1)} \quad (\text{AX-1.6})$$

where

$$w^{(1)} = L(x^{(1)}) = Kx^{(1)} - y$$

$$\langle Kw^{(n-1)}, w^{(n-1)} \rangle \neq 0 \quad \text{for } n = 1, 2, \dots$$

Now if $w^{(n-1)} = 0$ and $\langle Kw^{(n-1)}, w^{(n-1)} \rangle = 0$, then $x^{(n-1)}$ is the solution.

In this problem we assume that operator K is positive definite or more precisely that

$$m(x, x) \leq (Kx, x) \leq M(x, x) \quad (\text{AX-1.7})$$

where $0 < m \leq M < +\infty$

Under this condition the inverse operator A^{-1} exists and equation (AX-1.1) has the unique solution x^* then

$$Kx^* = y \quad \text{or} \quad x^* = K^{-1}y$$

As regards to the convergence of the process in order to obtain this solution and its speed of convergence we have the following theorem. The detailed discussion can be found in reference (39).

Theorem F: The successive approximation $x^0, x^1 \dots$ converge strongly (i.e., in norm) to the solution of equation (AX-1.1) with the speed of a geometrical progression if it can be proved that

$$\|x^n - x^*\|^2 \leq \frac{c}{m} \left(\frac{M-m}{M+m} \right)^{2n} \quad (\text{AX-1.8})$$

$$\text{as } n \rightarrow \infty, \quad x^n \rightarrow x^*$$

where $c = f(x^0) - f(x^*)$.

Corollary: The rapidity of convergence of the process obtained in theorem F was of the order of a geometric progression with ratio

$$q = \frac{M - m}{M + m} \quad (\text{AX-1.9})$$

We shall apply the above steepest descent method to the solution of Problem 3 in Section 2.4.1 in Chapter II.

Let $\{f_1\}_1^n$ be a basis for $M = N_{F^t}^\perp$

then $\{g_1\}_1^n$ defined by $(I + F^*F)g_1 = f_1$ is a basis for S_u

and $\{g_1, Fg_1\}_1^n$ is a basis for $Q = N_{G^t}^\perp$

where

$$S_u = \left\{ u : u = (I + F^*F)^{-1}v, \quad v \in M \right\}$$

$$Q = \left\{ (u, Fu) : (I + F^*F)u = v, \quad v \in M \right\},$$

and the bounded linear transformation $I + F^*F$ from S_u to M is nonsingular.

Given $f_1 \in M$ we can compute g_1 that satisfies $Kg_1 = f_1$, where $K = (I + F^*F)$. By the method of steepest descent the sequence of $g_1^{(n)}$ at n -th step is given by (AX-1.6). In the process of computation we let $g_1^{(n)}$ be an arbitrary element in $L_2(t_0, t_f)$. To evaluate (AX-1.4) in an efficient manner we first compute $W^{(0)}(s)$ given by

$$\begin{aligned} W^{(0)}(s) &= L(g_1^{(0)}) - Kg_1^{(0)}(s) - f_1(s) \\ &= g_1^{(0)}(s) + B^*(s) \Phi^*(t_0, s) \int_s^{t_f} \Phi^*(t, t_0) \left\{ \Phi(t, t_0) \right. \\ &\quad \left. \int_{t_0}^t \Phi(t_0, s_1) B(s_1) g_1^{(0)}(s_1) ds_1 \right\} dt - f_1(s) \end{aligned}$$

then

$$\langle L(g_1^{(0)}), L(g_1^{(0)}) \rangle = \langle W^{(0)}, W^{(0)} \rangle = \int_{t_0}^{t_f} |W^{(0)}(s)|^2 ds$$

and

$$\begin{aligned} \langle KL(x^{(0)}), L(x^{(0)}) \rangle &= \langle KW^{(0)}, W^{(0)} \rangle = \langle W^{(0)}, W^{(0)} \rangle + \langle FW^{(0)}, FW^{(0)} \rangle \\ &= \int_{t_0}^{t_f} \left\{ |W^{(0)}(s)|^2 + |P(s)|^2 \right\} ds, \end{aligned}$$

$$\text{where } P(s) = FW^{(0)} = \Phi(s, t_0) \int_{t_0}^s \Phi(t_0, \tau) B(\tau) W^{(0)}(\tau) d\tau$$

According to (AX-1.5) the iterate $g_1^{(1)}$ is then given by

$$g_1^{(1)}(t) = g_1^{(0)}(t) - \frac{\langle w^{(0)}, w^{(0)} \rangle}{\langle Kw^{(0)}, w^{(0)} \rangle} w^{(0)}(t) \quad i = 1, 2 \dots n$$

Having determined the basis of $\{g_i\}_1^n$ we are now in the position to find the optimal u^0 . Because u is an element in the space S_u , u^0 can be expressed as a linear combination of $\{g_i\}_1^n$ namely:

$$u^0 = \sum_{i=1}^n g_i c_i.$$

But $F^t u = x$, where $x = \sum_{i=1}^n \phi_i \beta_i$. By the same argument as

stated in Problem 1 in Section 2.2.1, the constant vector c satisfies the following equation:

$$\beta = G c \quad \text{or} \quad c = G^{-1} \beta$$

where

$$c = \text{col. } (c_1 \dots c_n)$$

$$\beta = \text{col. } (\beta_1 \dots \beta_n) \quad \beta_i = \sum_{i=1}^n \langle \phi_i^+, x \rangle$$

$$G = [\langle f_i, g_j \rangle]$$

AX-2: Minimization of $z=f(x)$ with the Inequality Constraint $g(x) \leq 0$

In the following analysis we shall seek a method for the minimization of $z = f(x)$ with the inequality constraint equation $g(x) \leq 0$ where $x = (x_1 \dots x_n)$.

First we shall give some important definitions and theorems in the following:

(a) Convex set: A set D of a space E is called convex if all segments joining any two points of the set D are

contained in this set, i.e., if $x \in D$ and $y \in D$ then

$$[x\theta + (1-\theta)y] \in D \quad \text{for } 0 \leq \theta \leq 1 \quad (\text{AX-2.1})$$

(b) Convex function: If D is a convex set of space E a real valued function defined in D is called a convex function if for x and y in D the following holds

$$f[(1-\theta)x + \theta y] \leq (1-\theta)f(x) + \theta f(y) \quad \text{for } 0 \leq \theta \leq 1 \quad (\text{AX-2.2})$$

(c) Absolute (global) maximum: A function $f(x)$ is said to have a absolute (global) maximum at x^* in D if the following inequality holds for any point of the domain D .

$$f(x) \leq f(x^*) \quad (\text{AX-2.3})$$

(d) Absolute (global) minimum: A function $f(x)$ is said to have a absolute (global) minimum at x^* in D if the following inequality holds for any point of the domain D .

$$f(x) \geq f(x^*) \quad (\text{AX-2.4})$$

(e) Relative or local maximum (minimum): $f(x)$ has a relative maximum (minimum) if $\epsilon > 0$ can be found such that f has an absolute maximum (minimum) at x^* in D which satisfy

$$|x_i - x_i^*| < \epsilon \quad \text{for } i = 1, 2, \dots, n$$

Theorem G: If $f(x)$, $x = (x_1 \dots x_n)$ is a twice continuously differentiable function in a convex domain D it is convex in D if and only if the Hessian matrix defined by

$$\mathcal{H} = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right] \quad i, j = 1 \dots n \quad (\text{AX-2.5})$$

is positive definite for every point in D .

Proof:

Necessity: From the extension of mean value theorem

$$f(x) = f(x_0) + \langle \nabla f(x_0), (x-x_0) \rangle + 1/2 \langle (x-x_0), \mathcal{H}_{\underline{x}}(x-x_0) \rangle$$

where $x_0 \ll \underline{x} \ll x$

For $0 < \theta < 1$ we write

$$\begin{aligned} & (1-\theta)f(x) + \theta f(y) - f[(1-\theta)x + \theta y] \\ &= (1-\theta) \left\{ f(x) - f[(1-\theta)x + \theta y] \right\} + \theta \left\{ f(y) - f[(1-\theta)x + \theta y] \right\} \end{aligned}$$

(AX-2.6)

Applying mean value theorem to the expansion of $f(x)$ and $f(y)$ equation (AX-2.6) reduces to

$$\begin{aligned} & (1-\theta)f(x) + \theta f(y) - f[(1-\theta)x + \theta y] \\ &= (1-\theta) \left\{ \langle \nabla f_{(1-\theta)x + \eta}, \theta x - \theta y \rangle + 1/2 \langle (\theta x - \theta y), \mathcal{H}_{\xi}(\theta x - \theta y) \rangle \right\} + \\ & \theta \left\{ \langle \nabla f_{(1-\theta)x + \theta y}, (1-\theta)(x-y) \rangle + 1/2 \langle (1-\theta)(x-y), \mathcal{H}_{\eta}(1-\theta)(x-y) \rangle \right\} \end{aligned}$$

(AX-2.7)

where $x \ll \xi \ll (1-\theta)x + \theta y$

and $y \ll \eta \ll (1-\theta)x + \theta y$

Since $\mathcal{H}$ is a positive definite matrix and the first and third term cancel, we have

$$(1-\theta)f(x) + \theta f(y) - f[(1-\theta)x + \theta y] \geq 0$$

from which $f(x)$ is convex.

Sufficient: If $\mathcal{H}$ is not a positive definite matrix at x_0 then there is a vector z such that

$$\langle z, \mathcal{H}(x_0) z \rangle < 0.$$

Therefore for all x in a neighborhood of x_0 , we also have

$$\langle z, \mathcal{H}(x) z \rangle < 0.$$

In particular, when $x = \xi$, and $y = \eta$ since the first expression in equation (AX-2.7) is negative this violates the convexity condition which is a contradiction.

Theorem H: If $f(x)$ and $g(x)$ are differentiable and convex, and define the set of points $R = \{x : g(x) \leq 0\}$, then:

(a) R is a convex set.

(b) Any local minimum of $f(x)$ is global minimum in R .

Proof:

(a) Let $x = (x_1 \dots x_n)$ and $y = (y_1 \dots y_n)$ be two points in R the $g(x) \leq 0$, $g(y) \leq 0$ and

$$g[(1-\theta)x + \theta y] \leq (1-\theta)g(x) + \theta g(y) \leq 0$$

where $0 < \theta \leq 1$, therefore $(1-\theta)x + \theta y \in R$.

(b) Let x^* and x be two points in R , where x^* is the value of x that gives the local minimum of $f(x)$. Now since $f(x)$ is a convex function, we have

$$(1-\theta)f(x) + \theta f(x^*) \geq f[(1-\theta)x + \theta x^*]$$

or

$$f(x) - f(x^*) \geq \frac{f[x^* + \theta(x - x^*)] - f(x^*)}{\theta} \quad . \quad 0 < \theta \leq 1$$

This relation will remain true as θ approaches to zero. By taking the limit $\theta \rightarrow 0$, we have

$$f(x) - f(x^*) \geq \langle \nabla f(x^*), (x - x^*) \rangle = 0 \quad \text{for all } x \in R,$$

since at extremum point the gradient is zero. By the definition of global minimum, it is clear that $f(x^*)$ is also a global minimum.

Let us first consider the problem of minimizing $f(x)$ with equality constraint $g(x)=0$. If x and z are two elements in the linear normed space X , and the functional

$f(x)$ and the constraint function $g(x)$ are defined on X , then the differential of functional $f(x)$ and $g(x)$ at the point x for the increment z have, respectively, the following form

$$df(x,z) = \langle \nabla f(x), z \rangle + \alpha_f(x,z) \simeq \langle \nabla f(x), z \rangle \quad (\text{AX-2.8})$$

$$dg(x,z) = \langle \nabla g(x), z \rangle + \alpha_g(x,z) \simeq \langle \nabla g(x), z \rangle = 0 \quad (\text{AX-2.9})$$

where α_f and α_g (denoted by α) are the higher order term in z in the sense that

$$\frac{\|\alpha(x,z)\|}{\|z\|} \rightarrow 0 \quad \text{for } \|z\| \rightarrow 0$$

If we assume $X = H$ a real Hilbert space, equation (AX-2.8) and (AX-2.9) imply that maximum change in f is attained precisely when z is chosen in the direction orthogonal to $\nabla g(x)$ or in the direction of the projection of $\nabla f(x)$ on the tangent plane to the surface $g(x) = 0$ at the point x . This follows from the fact that if ∇f is decomposed into two components one is parallel to ∇g designated by ∇f_1 and the other is orthogonal to ∇g designated by ∇f_2 , then we have

$$df(x,z) = \langle \nabla f_1(x) + \nabla f_2(x), z \rangle = \langle \nabla f_2(x), z \rangle \quad (\text{AX-2.10})$$

From the Schwarz inequality df has a most negative value (steepest descent) if z is chosen such that

$$z = - \frac{\nabla f_2(x)}{\|\nabla f_2(x)\|} \quad \text{for } \|z\| = 1$$

which is a unit vector along the projection of ∇f on the tangent plane to the surface $g(x) = 0$ at x . If $x^{(0)}$ is the initial estimation then $x^{(1)}$ which is adopted as next point in the iteration is given by

$$x^{(1)} = x^{(0)} - k_0 \frac{\nabla f_2(x^{(0)})}{\|\nabla f_2(x^{(0)})\|} \quad (\text{AX-2.11})$$

We can compute $\nabla f_2(x)$ in term of ∇f and ∇g by use of the fact that $\nabla f_2 = \nabla f - \nabla f_1 = \nabla f - c \nabla g$

$$\text{and} \quad \langle \nabla g, \nabla f_2 \rangle = 0$$

which gives

$$c = \frac{\langle \nabla g, \nabla f \rangle}{\langle \nabla g, \nabla g \rangle}$$

$$\text{Hence} \quad \nabla f_2 = \nabla f - \frac{\langle \nabla g, \nabla f \rangle}{\langle \nabla g, \nabla g \rangle} \nabla g = (1 - \nabla g \frac{\nabla g^*}{\|\nabla g\|^2}) \nabla f \quad (\text{AX-2.12})$$

where ∇g^* is the transpose of ∇g .

Lemma: Let $\nabla f \in R^n$ and $R^n = L \oplus L^\perp$ where ∇f has the unique representation $\nabla f = \nabla f_1 + \nabla f_2$ with $\nabla f_1 \in L^\perp$, $\nabla f_2 \in L$ then $(1 - \nabla g \frac{\nabla g^*}{\|\nabla g\|^2})$ is a projection operator, designated

by P_L , and defined on R^n with range in the subspace L , where $\nabla g \in L^\perp$.

$$\begin{aligned} \text{Proof: } P_L \nabla f_2 &= (1 - \nabla g \frac{\nabla g^*}{\|\nabla g\|^2}) \nabla f_2 \\ &= \nabla f_2 - \nabla g \frac{\langle \nabla g, \nabla f_2 \rangle}{\|\nabla g\|^2} \\ &= \nabla f_2 \quad \text{where } \nabla f_2 \in L \end{aligned}$$

and

$$P_L \nabla g = (1 - \nabla g \frac{\nabla g^*}{\|\nabla g\|^2}) \nabla g = \nabla g - \nabla g = 0$$

where $\nabla g \in L^\perp$.

Hence (AX-2.11) may be written in the form

$$x^{(j+1)} = x^{(j)} - k_j \frac{P_L \nabla f(x^{(j)})}{\|P_L \nabla f(x^{(j)})\|}$$

Unfortunately $x^{(j+1)}$ determined in this way does not lie on the surface $g(x)=0$ because of the omission of $\partial_g(x,z)$ in equation (AX-2.9) which leads to the property that z is chosen in the direction along the projection of ∇f on the tangent plane to the surface $g(x)=0$. Strictly we shall find the projection of ∇f on the surface of $g(x)=0$. Let $P_S \nabla f$ be the chord formed by the shadow of ∇f on the surface $g(x)=0$ (see Figures AX-2.1 and AX-2.2) then the corrected iteration is given by

$$x^{(j+1)} = x^{(j)} - k_j \frac{P_S \nabla f(x^{(j)})}{\|P_S \nabla f(x^{(j)})\|} \quad (\text{AX-2.13})$$

where $P_S \nabla f(x^{(j)}) \triangleq [\nabla f(x^{(j)}) - e_1 \nabla g(x^{(j)})]$ (AX-2.14)

The scalar factor e_1 can be determined by substituting $x^{(j+1)}$ in $g(x^{(j+1)})=0$. The scalar factor k_j can be selected by forming a Taylor series about the point $x^{(j)}$ (taking first three terms in the expansion):

$$f(x^{(j+1)}) = f(x^{(j)}) + k_j \langle \nabla f(x^{(j)}), z^{(j)} \rangle + \frac{k_j^2}{2} \langle z^{(j)}, \mathcal{H}(x^{(j)}) z^{(j)} \rangle$$

For $f(x^{(j+1)})$ to be a minimum, we have

$$k_j = - \frac{\langle \nabla f(x^{(j)}), z^{(j)} \rangle}{\langle z^{(j)}, \mathcal{H}(x^{(j)}) z^{(j)} \rangle}$$

where
$$z^{(j)} = \frac{P_S \nabla f(x^{(j)})}{\|P_S \nabla f(x^{(j)})\|}$$

Substituting the above k_j in (AX-2.13) we have

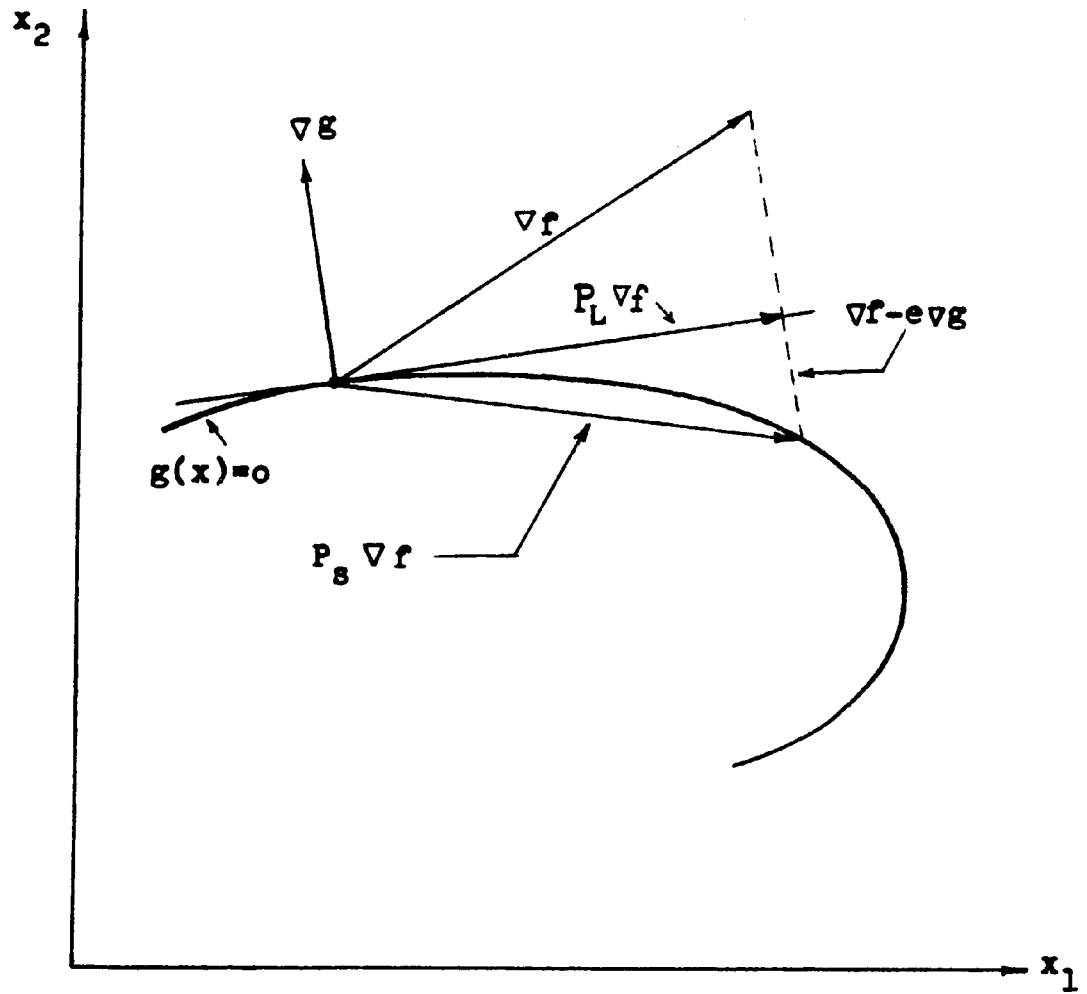


Fig. AX-2.1.--A geometric meaning of P_L and P_S .

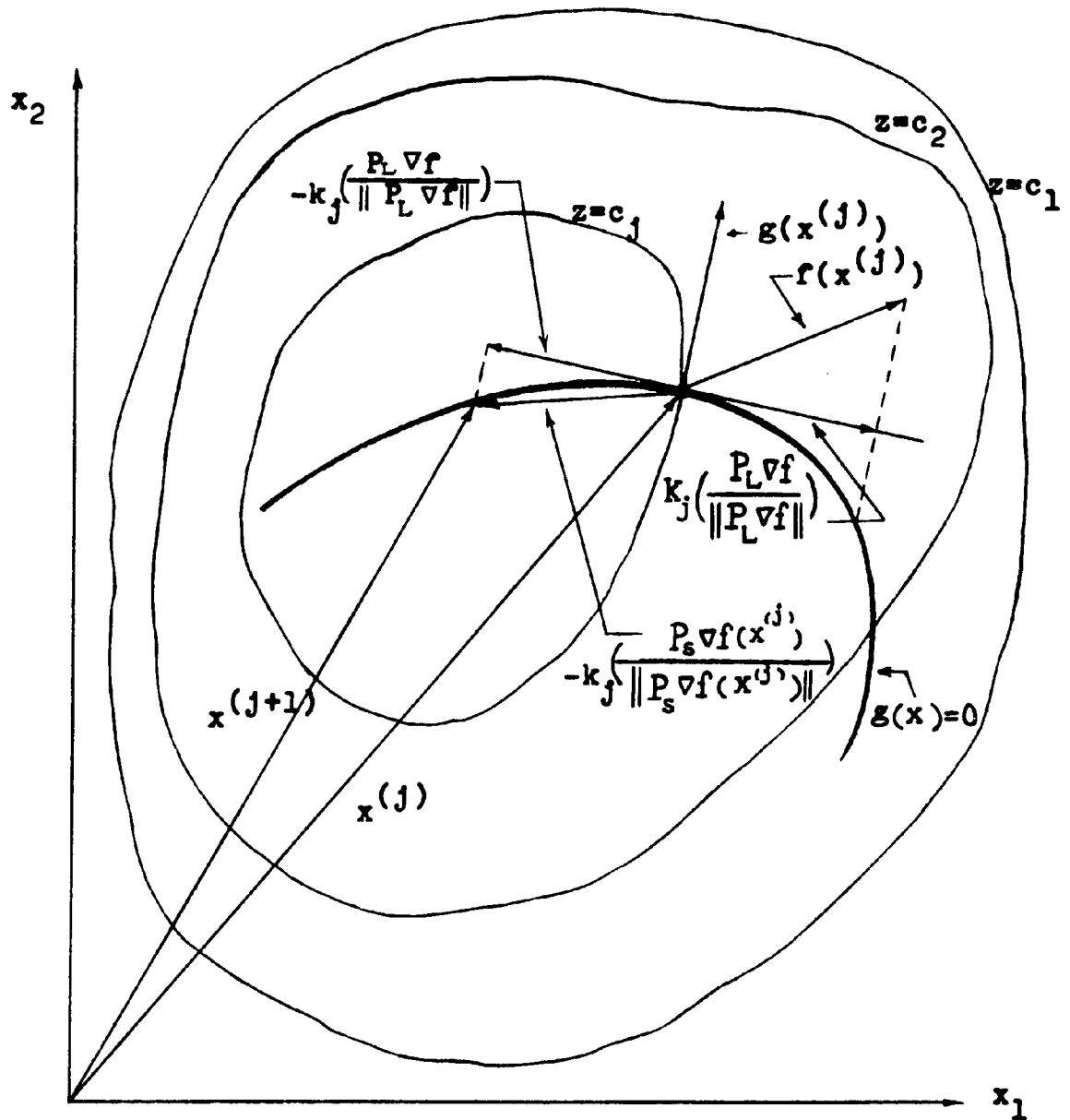


Fig. AX-2.2.--A geometric interpretation of the method of steepest descent in the minimization of $z = f(x_1, x_2)$ with constraint $g(x_1, x_2) = 0$.

$$x^{(j+1)} = x^{(j)} - \frac{\langle \nabla f(x^{(j)}), z^{(j)} \rangle}{\langle z^{(j)}, \nabla f(x^{(j)}) \rangle} z^{(j)} \quad (\text{AX-2.15})$$

Now we shall consider the solution of minimization of $f(x)$ with inequality constraint $g(x) \leq 0$ in the following.

Let the feasible region (i.e., one which satisfies the constraints $g(x) \leq 0$) be denoted by $R = \{x : g(x) \leq 0\}$. Starting with any point $x^{(0)}$ inside the region R we move along the vector $-\nabla f(x^{(0)})$ a distance equal to $k_0 z^{(0)}$ to get $x^{(1)}$ expressed by

$$x^{(1)} = x^{(0)} - k_0 z^{(0)}$$

$$\text{where } k_0 = \frac{\langle \nabla f(x^{(0)}), z^{(0)} \rangle}{\langle z^{(0)}, \nabla f(x^{(0)}) \rangle}$$

$$\text{and } z^{(0)} = \frac{\nabla f(x^{(0)})}{\|\nabla f(x^{(0)})\|}$$

if $x^{(1)}$ is inside region R and $f(x^{(1)}) < f(x^{(0)})$ then $x^{(1)}$ is the new point and this process is repeated to find $x^{(2)}, x^{(3)}, \dots$

If $x^{(1)}$ lies on the boundary of R and $f(x^{(1)}) < f(x^{(0)})$ then we shall apply projection gradient method to find $x^{(2)}$. Therefore in general we may use the following gradient criterion to seek the new point $x^{(j+1)}$ by starting from $x^{(j)}$ such that

$$x^{(j+1)} = x^{(j)} - k_j z^{(j)} \quad (\text{AX-2.16})$$

with

$$k_j = \frac{\langle \nabla f(x^{(j)}), z^{(j)} \rangle}{\langle z^{(j)}, \nabla f(x^{(j)}) \rangle} \quad (\text{AX-2.17})$$

where

$$\begin{aligned}
 z^{(j)} &= \frac{\nabla f(x^{(j)})}{\|\nabla f(x^{(j)})\|} \quad \text{when } x^{(j+1)} \text{ lies inside } \mathcal{R}, \text{ i.e., } g(x^{(j+1)}) < 0 \\
 &= \frac{P_S \nabla f(x^{(j)})}{\|P_S \nabla f(x^{(j)})\|} \quad \text{when } x^{(j+1)} \text{ is on the boundary of } \mathcal{R}, \\
 &\quad \text{i.e., } g(x^{(j+1)}) = 0 \quad (\text{AX-2.18})
 \end{aligned}$$

If $x^{(1)}$ does not belong to $\mathcal{R}$ then we must calculate the intersection of $g(x)=0$ with the line passing through $x^{(0)}$ and $x^{(1)}$. The point of intersection will be selected as a new point which is on the boundary of $\mathcal{R}$, i.e., $g(x)=0$.

Theorem I: Let $f(x)$ and $g(x)$ be differentiable and convex. Then the iterations given by (AX-2.16) converge to x^* which gives the minimum point of $f(x)$ with $g(x) \leq 0$.

Proof: From Taylor series expansion

$$\begin{aligned}
 f(x^{(j+1)}) - f(x^{(j)}) &\simeq k_j \langle \nabla f(x^{(j)}), z^{(j)} \rangle + 1/2 k_j^2 \\
 &\quad \langle z^{(j)}, \mathcal{A}(x^{(j)}) z^{(j)} \rangle
 \end{aligned}$$

and the expression of k_j in (AX-2.17) we have

$$\begin{aligned}
 f(x^{(j+1)}) - f(x^{(j)}) &\simeq - \frac{\langle \nabla f(x^{(j)}), z^{(j)} \rangle^2}{\langle z^{(j)}, \mathcal{A}(x^{(j)}) z^{(j)} \rangle} \\
 &\quad + \frac{1}{2} \frac{\langle \nabla f(x^{(j)}), z^{(j)} \rangle^2}{\langle z^{(j)}, \mathcal{A}(x^{(j)}) z^{(j)} \rangle} \\
 &= - 1/2 \frac{\langle \nabla f(x^{(j)}), z^{(j)} \rangle^2}{\langle z^{(j)}, \mathcal{A}(x^{(j)}) z^{(j)} \rangle} \quad (\text{AX-2.19})
 \end{aligned}$$

where

$$z^{(j)} = - \frac{\nabla f(x^{(j)})}{\|\nabla f(x^{(j)})\|} \quad \text{for } x^{(j)} \text{ inside } \mathcal{R}$$

$$= - \frac{P_S \nabla f(x^{(j)})}{\|P_S \nabla f(x^{(j)})\|} \quad \text{for } x^{(j)} \text{ on the boundary of } \mathcal{R}$$

Since f is convex, $\mathcal{H}$ is positive definite and we have $f(x^{(j+1)}) - f(x^{(j)}) < 0$. Now

$$f(x) \doteq f(x^{(j)}) + \langle \nabla f(x^{(j)}), x - x^{(j)} \rangle + 1/2 \langle x - x^{(j)}, \mathcal{H}(x^{(j)})(x - x^{(j)}) \rangle \quad (\text{AX-2.20})$$

Taking partial differential with respect to x , it gives

$$\nabla f(x) \doteq \nabla f(x^{(j)}) + \mathcal{H}(x^{(j)})(x - x^{(j)})$$

and then replacing $x^{(j)}$ by x^* and x by $x^{(j)}$, results in

$$f(x^{(j)}) \doteq \mathcal{H}(x^*)(x^{(j)} - x^*)$$

or

$$(x^{(j)} - x^*) \doteq \mathcal{H}(x^*)^{-1} \nabla f(x^{(j)}) \quad (\text{AX-2.21})$$

From (AX-2.20) we have

$$f(x^{(j)}) - f(x^*) \doteq 1/2 \langle x^{(j)} - x^*, \mathcal{H}(x^*)(x^{(j)} - x^*) \rangle \quad (\text{AX-2.22})$$

Note that $\nabla f(x^*) = 0$, and $\mathcal{H}^{-1}$ exists because $\mathcal{H}$ is a positive definite matrix.

Therefore from (AX-2.19), (AX-2.21) and (AX-2.22)

we have:

$$\frac{f(x^{(j)}) - f(x^{(j+1)})}{f(x^{(j)}) - f(x^*)} \doteq \frac{\langle \nabla f(x^{(j)}), z^{(j)} \rangle^2}{\langle z^{(j)}, \mathcal{H}(x^{(j)})z^{(j)} \rangle \langle \nabla f(x^{(j)}), \mathcal{H}^{-1}(x^*)\nabla f(x^{(j)}) \rangle} \quad (\text{AX-2.23})$$

Let m and M denote the smallest and largest eigenvalues respectively of $\mathcal{H}(x)$ and for all $x \in \mathcal{R}$. Then for every vector z we have

$$m \langle z, z \rangle \leq \langle z, \mathcal{A} z \rangle \leq M \langle z, z \rangle$$

$$\text{and } M^{-1} \langle z, z \rangle \leq \langle z, \mathcal{A}^{-1} z \rangle \leq m^{-1} \langle z, z \rangle$$

Hence

$$\frac{f(x^{(j)}) - f(x^{(j+1)})}{f(x^{(j)}) - f(x^*)} \leq \frac{m \langle \nabla f(x^{(j)}), z^{(j)} \rangle^2}{M \langle z^{(j)}, z^{(j)} \rangle \langle \nabla f(x^{(j)}), \nabla f(x^{(j)}) \rangle} \quad (\text{AX-2.24})$$

where $\leq$ means greater than or approximately equal to.

Recalling that

$$z^{(j)} \begin{cases} = - \frac{\nabla f(x^{(j)})}{\|\nabla f(x^{(j)})\|} & \text{when } x \text{ lies inside } \mathcal{R} \\ = - \frac{P_s \nabla f(x^{(j)})}{\|P_s \nabla f(x^{(j)})\|} & \text{where } x \text{ is on the boundary of } \mathcal{R}. \end{cases}$$

Equation (AX-2.24) reduces to the form

$$\frac{f(x^{(j)}) - f(x^{(j+1)})}{f(x^{(j)}) - f(x^*)} \leq \frac{m}{M} \quad \text{when } x \text{ lies inside } \mathcal{R}$$

and

$$\begin{aligned} \frac{f(x^{(j)}) - f(x^{(j+1)})}{f(x^{(j)}) - f(x^*)} &\leq \frac{m \langle \nabla f(x^{(j)}), P_s \nabla f(x^{(j)}) \rangle^2}{M \langle P_s \nabla f(x^{(j)}), P_s \nabla f(x^{(j)}) \rangle \langle \nabla f(x^{(j)}), \nabla f(x^{(j)}) \rangle} \\ &\leq \frac{m}{M} \end{aligned}$$

when x is on the boundary of $\mathcal{R}$ since from the Schwarz inequality we have $\|A\|^2 \|B\|^2 \geq \langle A, B \rangle^2$. Therefore, in any case, we have

$$\frac{f(x^{(j)}) - f(x^{(j+1)})}{f(x^{(j)}) - f(x^*)} \leq \frac{m}{M} \quad \text{for } x \in \mathcal{R} \quad (\text{AX-2.25})$$

If we subtract each side of equation (AX-2.25) from unity we have

$$f(x^{(j+1)}) - f(x^*) \leq (1 - \frac{m}{M}) [f(x^{(j)}) - f(x^*)], \quad j=0,1,2,\dots$$

(AX-2.26)

Consequently

$$f(x^{(j)}) - f(x^*) \leq (1 - \frac{m}{M})^j [f(x^{(0)}) - f(x^*)]$$

But

$$f(x^{(j)}) - f(x^*) \geq 1/2 \langle (x^{(j)} - x^*), \nabla f(x^*)(x^{(j)} - x^*) \rangle \geq \frac{m}{2} \langle (x^{(j)} - x^*), (x^{(j)} - x^*) \rangle$$

Hence $\|x^{(j)} - x^*\| \leq \frac{2}{m} (1 - \frac{m}{M})^j [f(x^{(0)}) - f(x^*)]$

as $j \rightarrow \infty$, $\|x^{(j)} - x^*\| \rightarrow 0$

Remarks: (1) If $f(x)$ and $g(x)$ are differential and convex we can always find the unique minimum of $f(x)$ in the set $\mathcal{R}$ by the method of steepest descent and the rapidity of convergence of the process obtained in theorem I was of order of a geometric progression with ratio $1 - \frac{m}{M}$.

(2) In the event that the perpendicular line at the tip of the vector

$$-k_j \frac{P_L \nabla f(x^{(j)})}{\|P_L \nabla f(x^{(j)})\|}$$

to the tangent plane at $x=x^{(j)}$ does not intersect the surface $g(x)=0$ we may cut down the size of step k_j such that the new point $x^{(j+1)} \in \mathcal{R}$.

APPENDIX AXI

A PROOF OF $H(F)^\perp = VH(F^*)$

Let (x, y) be an arbitrary element of $H(F)^\perp$, then for every $u \in H_1$

$$\begin{aligned} 0 &= \langle (x, y), (u, Fu) \rangle_{H(F)} = \langle x, u \rangle_{H_1} + \langle y, Fu \rangle_{H_2} = \\ &\langle x, u \rangle_{H_1} + \langle F^*y, u \rangle_{H_1} = \langle x + F^*y, u \rangle_{H_1} \end{aligned} \quad (\text{AXI-1})$$

which implies $x = -F^*y$ or that (x, y) is of the form

$$(x, y) = (-F^*y, y) = V(y, F^*y) \quad (\text{AXI-2})$$

thus $(x, y) \in VH(F^*)$ and hence $H(F)^\perp \subset VH(F^*)$.

Conversely if $(x, y) \in VH(F^*)$ then

$$(x, y) = V(y, F^*y) = (-F^*y, y) \quad (\text{AXI-3})$$

For any (u, Fu) however $(u, Fu) \perp (-F^*y, y)$ since

$$\begin{aligned} \langle (u, Fu), (-F^*y, y) \rangle_{H(F)} &= \langle u, -F^*y \rangle_{H_1} + \langle Fu, y \rangle_{H_2} = \\ &-\langle Fu, y \rangle_{H_2} + \langle Fu, y \rangle_{H_2} = 0 \end{aligned}$$

hence $(-F^*y, y) \in H(F)^\perp$ which implies $VH(F^*) \subset H(F)^\perp$ and the lemma is proved.

APPENDIX AXII

A PROOF OF LEMMA (1) IN SECTION 2.4.1

Property (1) follows directly from the definition.

That the function $\langle (f_1, g_1), (f_2, g_2) \rangle_{H(F)} = \langle f_1, f_2 \rangle_{H_1} + \langle g_1, g_2 \rangle_{H_2}$

is an inner product on $H(F)$ is also easily shown. For instance, if $W_1 = (f_1, g_1)$ and $W_2 = (f_2, g_2)$ then

$$\begin{aligned} \langle W_1, W_2 \rangle_{H(F)} &= \langle f_1, f_2 \rangle_{H_1} + \langle g_1, g_2 \rangle_{H_2} = \overline{\langle f_2, f_1 \rangle_{H_1}} + \\ &\quad \overline{\langle g_2, g_1 \rangle_{H_2}} = \overline{\langle W_2, W_1 \rangle_{H(F)}} \end{aligned}$$

where $\bar{W}$ denotes the complex conjugate of W . Thus it remains only to show that $H(F)$ is complete and closed when F is bounded. Let $\{W_n = (f_n, g_n)\}$ denote a Cauchy sequence in $H(F)$ then for arbitrary ϵ an N exists such that

$$\|W_m - W_n\|_{H(F)}^2 = \|f_m - f_n\|_{H_1}^2 + \|g_m - g_n\|_{H_2}^2 < \epsilon \quad \text{for } m, n > N. \quad \text{Clearly}$$

this implies both $\|f_m - f_n\|_{H_1}^2 < \epsilon_1$ and $\|g_m - g_n\|_{H_2}^2 < \epsilon_2$, $\epsilon_1 + \epsilon_2 = \epsilon$,

for $m, n > N$. Since H_1 and H_2 are both complete the sequence $\{f_n\}$ and $\{g_n\}$ both converge to points f and g of those spaces. Since $g_n = Ff_n$ for all n it follows that $g = Ff$ and hence the element w which is the limit of the sequence $\{W_n\}$ is given by $W = (f, g) \in H(F)$ and hence $H(F)$ is complete.

Let $(u_1, Fu_1), (u_2, Fu_2) \dots$ be the sequence of points of $H(F)$ since F is a bounded linear operator this implies

that F is continuous, therefore

$$Fu_n \rightarrow Fu \text{ as } u_n \rightarrow u$$

This leads to $\lim_{n \rightarrow \infty} (u_n, Fu_n) = (u, Fu) \in H(F) \subset H_1 \times H_2$.

We must show that $H(F)$ is closed or equivalently that its complement, denoted by $H(F)^c$, is open. If $H(F)^c$ is not open then there must be a point $(c, Fc) \in H(F)^c$ such that for each neighborhood N of (c, Fc)

$$N \not\subset H(F)^c$$

In particular the sphere $s((c, Fc), \frac{1}{n}) \not\subset H(F)^c$ for $n = 1, 2, 3, \dots$. Thus there is a sequence of points $(u_1, Fu_1), (u_2, Fu_2), \dots$ such that $(u_n, Fu_n) \in H(F)$ and $\lim_{n \rightarrow \infty} (u_n, Fu_n) = (c, Fc) \in H(F)$ which contradicts to the assumption that $(c, Fc) \in H(F)^c$ therefore $H(F)^c$ is open and $H(F)$ is closed.

LIST OF REFERENCES

1. Akhiezer, N. I. and Glazman, I. M. Theory of Linear Operators in Hilbert Space. Frederick Ungar Publishing Co., New York, 1961.
2. Athans, M. and Falb, P. L. "Time Optimal Control for a Class of Nonlinear System," IEEE Transactions on Automatic Control. October 1963, vol. AC-8, No. 4.
3. Athans, M., Falb, P. L., and Lacoss, R. T. "On Optimal Control of Self-Adjoint Systems," to be presented at 1963 Joint Automatic Conference. Minneapolis, Minn., June 19-21, 1963.
4. Athans, M. "Optimal Control for Linear Time Invariant Plant with Time, Fuel and Energy Constraints," IEEE Transactions on Applications and Industry, January 1963, Vol. 81, pp. 321-325.
5. Athans, M. "Minimum Fuel Feedback Control Systems: Second Order Case," IEEE Transactions on Applications and Industry. March 1963, Vol. 82, pp. 8-17.
6. Athans, M., Falb, P. L., and Lacoss, R. T. "Time Optimal Velocity Control of a Spinning Space Body," IEEE Paper No. 63-98.
7. Athans, M., Falb, P. L. and Lacoss, R. T. "Time, Fuel and Energy Optimal Control of Nonlinear Norm-Invariant Systems," IEEE Transactions on Automatic Control. July 1963, Vol. AC-8, No. 3.
8. Barbashin, E. A. "Estimating the Mean Square Deviation from a Given Trajectory," Avtomatika i Telemekhanik. July 1960, Vol. 21, No. 7, pp. 941-950.
9. Balakrishnan, A. V. "An Operator Theoretic Formulation of a Class of Control Problems and a Steepest Descent Method of Solution," Journal of Industrial and Applied Mathematics on Control, 1963, Vol. 1, No. 2, pp. 109-127.
10. Bellman, R. Dynamic Programming, Princeton University Press, Princeton, N.J., 1957.
11. Bellman R., Glicksberg, I. and Gross, O. "On the Bang Bang Control Problem," Quarterly of Applied Mathematics. 1956, Vol. 14, pp. 11-18.

12. Bendat, J. S. Principles and Applications of Random Noise Theory. John Wiley and Sons, Inc., 1958.
13. Bliss, G. A. Calculus of Variation. Open Court Publishing Co., Chicago, 1925.
14. Boltyanskii, V. G., Gamkrelidze, R. V. and Pontryagin, L. S. "On the Theory of Optimal Processes," Doklady Akad. Nauk SSSR. 1956, Vol. 110, No. 1.
15. Boltyanskii, R. V. "The Maximum Principle in Optimal Process Theory," Doklady Akad. Nauk SSSR. 1958, Vol. 119, No. 6.
16. Booton, R. C. "Optimal Design of Final Value Control Systems," Proceedings of the Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, New York, 1956, pp. 233-242.
17. Bushaw, D. W. E.T.T. Report 469. Stevens Institute of Technology, Hoboken, New Jersey, 1953.
18. Calabro, S. Reliability Principles and Practices. McGraw-Hill, New York, 1962.
19. Chang, S. S. L. Synthesis of Optimum Control Systems. McGraw-Hill Book Company, Inc., 1961.
20. Coddington, E. A. and Levinson, N. T. Theory of Ordinary Differential Equations. McGraw-Hill, 1955.
21. Collina, G. L. and Dorato, P. "Application of Pontryagin's Maximum Principle: Linear Control Systems," Research Report No. PIBMRI-1015-62 for Office of Scientific Research, Washington, D.C., June 1962.
22. Cuba, R. E. and Kazda, L. F. "A Phase-Space Method for the Synthesis of Nonlinear Servomechanisms," AIEE Transactions on Applications and Industry, November 1956, Vol. 75, pp. 282-290.
23. Davenport, W. B. and Root, W. L. Random Signals and Noise. McGraw-Hill Book Company, 1958.
24. Desoer, C. A. "The Bang-Bang Servo Problem Treated by Variational Techniques," Information and Control. December 1959, Vol. 2, pp. 333-348.
25. Desoer, C. A. "Pontryagin's Maximum Principle and the Principle of Optimality," J. of Franklin Institute, May 1961.

26. Draper, C. S. and Li, Y. T. "Principles of Optimizing Control Systems and an Application to the Internal Combustion Engine," American Society of Mechanical Engineers, New York, 1951.
27. Eggleston, H. G. Convexity. Cambridge University Press, 1958.
28. Elsgole, L. E. Calculus of Variation. Addison-Wesley Publishing Company, Inc., 1962.
29. Fischbach, J. W. "Some Applications of Gradient Methods," Proceedings of Symposia in Applied Mathematics, the American Mathematical Society, 1956, Vol. VI, pp. 59-72.
30. Friedland, B. "The Structure of Optimum Control Systems," Transactions ASME. March 1962, pp. 1-10.
31. Fuller, A. T. "Phase Space in the Theory of Optimal Control," J. Electronics and Control. 1960, Vol. 8, pp. 381-400.
32. Gamkrelidze, R. V. "On the General Theory of Optimal Processes," Doklady Akad. Nauk SSSR. 1958, Vol. 123, No. 2, pp. 223-226.
33. Hoffman, K. and Kunze, R. Linear Algebra. Prentice-Hall, 1961.
34. Hopkin, A. M. "A Phase-Plane Approach to the Design of Saturating Servomechanism," Trans. AIEE Communication and Electronics. 1951, Vol. 70, pp. 631-639.
35. Hsieh, H. C. "Synthesis of Optimum Multivariable Control Systems by the Method of Steepest Descent," IEEE Transactions paper 63-134, February 1963.
36. Ito, Kiyoshi. "On Stochastic Equation," American Mathematical Society. 1951, Memoirs No. 4.
37. Kalman, R. E. "Contributions to the Theory of Optimal Control," Bol. Soc. Mat., Mexicana. 1961.
38. Kalman, R. E. "The Theory of Optimal Control and the Calculus of Variations," RIAS, Baltimore, Md., Report 61-3, 1961.
39. Kantarovich, L. V. "Functional Analysis and Applied Mathematics," Usp. Mat. Nauk., 3, 1948, pp. 89-185; English Translation, National Bureau of Standards, 1953.

40. Kao, R. C. "A Geometric Interpretation of Lagrange Multipliers," Tech. Report Rand-P-2713, c.3, AD 298197, February 1963.
41. Kaplan, W. Ordinary Differential Equations. Addison-Wilsey Publishing Company, Inc., Cambridge, Mass., 1961.
42. Kaplan, W. Advanced Calculus. Addison-Wesley Publishing Company, Inc., 1959.
43. Kelendzheridge, P. L. "On the Theory of Optimal Pursuit," Doklady Akad. Nauk SSSR. 1961, Vol. 138, pp. 529-532, English translation in Soviet Math.-Doklady, 1961, Vol. 2, pp. 654-656.
44. Kolmogorov, A. N. and Fomin, S. V. Functional Analysis. Graylock Press, Rochester, N.Y., 1957.
45. Kopp, R. E. "Pontryagin Maximum Principle," Optimization Techniques, Edited by Leitmann, G., Academic Press, New York, New York, 1962.
46. Krasovskii, N. N. "On the Theory of Optimum Regulation," Automation and Remote Control. November 1957, Vol. 18, pp. 1005-1016.
47. Krasovskii, N. N. "On the Theory of Optimal Control," Avtomat. i. Telemekh., 1957, Vol. 18, pp. 960-970.
48. Kuo, B. C. Automatic Control Systems. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1962.
49. Lanning, J. H. and Battin, R. W. Random Process in Automatic Control. McGraw-Hill, 1956.
50. Lasalle, J. P. Goodyear Aircraft Report, Akron, Ohio, 1953.
51. Lasalle, J. P. "The Time Optimal Control Problem," Contribution to the Theory of Nonlinear Oscillations, Princeton University Press, 1960, Vol. 5, pp. 1-24.
52. Lasalle J. P. "Time Optimal Control Systems," Proceeding of National Academic Science. April 1957, Vol. 45, pp. 573-577.
53. Lasalle, J. P. and Lefschetz, S. Stability by Liapunov's Direct Method with Applications. Academic Press, 1961.
54. Lee, E. B. "Mathematical Aspect of the Synthesis of Linear Minimum Response Time Controllers," IRE Transactions on Automatic Control, September 1960, Vol. AC-5, pp. 283-289.

55. Lee, E. B. "Design of Optimum Multi-variable Control Systems," J. Basic Eng. (Trans. ASME), 1961, Vol. 33, pp. 85-90.
56. Leitmann, G., Optimization Techniques with Applications to Aerospace Systems, Academic Press, Inc., New York, 1962.
57. Liusternik, L. A. and Sobolev, V. J. Elements of Functional Analysis, Frederick Ungar Publishing Co., New York, 1961.
58. McDonald, D. C. "Nonlinear Techniques for Improving Servo Performance," Proceedings of the National Electronic Conference, 1950, Vol. 6, pp. 400-421.
59. Meditch, J. S. and Neustadt, L. W. "An Application of Optimal Control of Midcourse Guidance," The Aerospace Corp., Los Angeles, Calif., Report A62-1732.2-5; July 1962.
60. Meditch, J. S. "Synthesis of a Class of Linear Feedback Minimum Energy Controls," IEEE Transactions on Automatic Control, Vo. AC-8, No. October 1963.
61. Mendelson, B., Introduction to Topology, Allyn and Bacon, Inc., Boston, 1962.
62. Merriam, C. W., III. "A Class of Optimum Control Systems," Journal of the Franklin Institute. April 1959, Vol. 267, pp. 267-281.
63. Natanson, I. P. Theory of Functions of Real Variable. Frederick Ungar Publishing Co., New York, 1961.
64. Neustadt, L. W. "Synthesizing Time Optimal Controls," Journal of Mathematic Analysis and Application, 1960, Vol. 1, No. 3, pp. 484-493.
65. Neustadt, L. W. "Time Optimal System with Position and Integral Limits," Journal of Mathematic Analysis and Applications, December 1961, Vol. 3, pp. 406-427.
66. Neustadt, L. W. "Minimum Effort Control Systems," J. SIAM Control, Ser. A, 1962, Vol. 1, No. 1.
67. Pearson, A. E. "Synthesis of a Minimum Energy Controller Subject to an Average Power Constraint," Joint Automatic Control Conference, New York, N.Y., June 27-29, 1962.

68. Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V. and Mishchenko, E. F. The Mathematical Theory of Optimal Processes, John Wiley and Sons, Inc., 1962.
69. Porter, W. A. "A New Approach to the General Minimum Energy Problem," Joint Automatic Control Conference. Stanford, California, 1964.
70. Ragazzini, J. R. and Franklin, G. F. Sampled-Data Control Systems. McGraw-Hill Book Company, Inc., New York, 1958.
71. Riesz, F. and Nagy, Sz. Functional Analysis. Frederick Ungar Publishing Co., New York, 1955.
72. Rosenbloom, A. "Final Value System with Total Effort Constraints," Proceedings of the First International Congress of the International Federation of Automatic Control. Butterworths, London, 1961.
73. Rosenbloom, P. C. "The Method of Steepest Descent," Proc. 7th AMS Symposium in Applied Mathematics, 1956, Vol. 6, pp. 127-176.
74. Rose, N. J. E.T.T. Report 495, Stevens Inst. of Technology, Hoboken, New Jersey, 1953.
75. Rozonoer, L. I. "L. S. Pontryagin's Maximum Principle in the Theory of Optimum Systems," Automation and Remote Control. October, November, December, 1959, Vol. 20, pp. 1288-1302, 1405-1421, 1517-1532.
76. Saaty, T. L. and Bram, J. Nonlinear Mathematics. McGraw-Hill Book Company, 1964.
77. Sammuel, J. C. and Eringen, A. C. "On Stochastic Linear System," Journal of Mathematics and Physics. July 1959, Vol. 38, No. 2, pp. 83-103.
78. Shilov, G. E. Introduction to the Theory of Linear Spaces. Prentice-Hall, Inc., New Jersey, 1961.
79. Simmon, G. F. Topology and Modern Analysis. McGraw-Hill Book Company, Inc., 1963.
80. Smith, O. J. M. Feedback Control Systems. McGraw-Hill Book Co., Inc., New York, N.Y., 1958.
81. Thrall, R. M. and Tornheim, L. Vector Spaces and Matrices. John Wiley and Sons, Inc., 1957.
82. Tou, J. T. Digital and Sampled-Data Control Systems. McGraw-Hill Book Company, Inc., New York, 1959.

83. Tsien, H. S. Engineering Cybernetics. McGraw-Hill, New York, 1954.
84. Zadeh, L. A. and Desoer, C. A. Linear System Theory The State Space Approach. McGraw-Hill Book Co., Inc., New York, 1963.