STANFORD UNIVERSITY

CENTER FOR SYSTEMS RESEARCH

Optimal Inputs for System Identification

by

Donald B. Reid
Department of Aeronautics and Astronautics

May 1972

Guidance and Control Laboratory

SUDAAR NO. 440

This research was sponsored by the
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
under Research Grants:
NASA NGR 05-020-526,
NASA NGL 05-020-007

OPTIMAL INPUTS FOR SYSTEM IDENTIFICATION

by

Donald B. Reid

Department of Aeronautics and Astronautics
Guidance and Control Laboratory
STANFORD UNIVERSITY
Stanford, California 94305

This research was sponsored by the
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
under Research Grants:
NASA NGR 05-020-526,
NASA NGL 05-020-007

SUDAAR NO. 440

May 1972

PRECEDING PAGE BLANK NOT FILMED

ABSTRACT

This thesis is concerned with determining optimal inputs to identify parameters of linear dynamic systems. Identification criteria are presented for linear dynamic systems with and without process noise. With process noise, the state equations are replaced by the Kalman filter equations. If the identification performance index is expanded in a Taylor's series with respect to the parameters to be identified, then maximizing the weighting factor of the quadratic term with respect to the inputs will insure that an identification algorithm will converge more rapidly and to a more accurate result than with non-optimal inputs. The expectation of this weighting factor is known as the Fisher information matrix, and its inverse is a lower bound for the covariance of the parameters. Direct and indirect methods of calculating the information matrix are presented for systems with and without process noise. The input design criterion used is the trace of the inverse of the information matrix. Minimizing this criterion appears to have some advantages over maximizing the trace of the information matrix.

With amplitude constraints on the input, the optimal input is full on in one direction or full on in the other direction (bang-bang). A gradient method is then used to minimize with respect to the switch times. The method is then applied to some simple illustrative examples. For sufficiently long tests, the optimal switch times are equally spaced and may be computed using the first few terms of the Fourier series for a square wave, minimizing with respect to the fundamental frequency. For reasonable amounts of deterministic input, the overall effect of process noise is to decrease the identification accuracy.

The method is then applied to finding the optimal elevator deflection to identify two damping derivatives of the short period longitudinal

Preceding page blank

equations of motion of an airplane. A simulation verifies the improvements of the optimal input over non-optimal inputs. Preliminary results are also obtained using the method to find the optimal aileron and rudder inputs to identify four damping derivatives of the lateral equations of motion of an airplane.

ACKNOWLEDGMENTS

I wish to thank my advisor, Professor Arthur E. Bryson, Jr., for his support and encouragement during the course of this research. I also wish to thank him for editing and helping me in the preparation of this dissertation.

I would like to thank Professors John V. Breakwell and J. David Powell for their reading of this manuscript and their helpful suggestions.

I also appreciate discussions with Dr. Raman Mehra and Dr. Dallas Denery on the subject of Identification.

I particularly wish to thank my wife, Marjorie, for her support and encouragement during this program, and for her typing of the first draft of this dissertation.

I also wish to thank Ida M. Lee for her careful typing of the final draft of this dissertation.

Finally, I would like to acknowledge the financial support of the National Aeronautics and Space Administration.

TABLE OF CONTENTS

TABLE OF CONTENTS (Contd)

(Cont)

TABLE OF CONTENTS (Contd)

TABLE OF CONTENTS (Contd)

List of Figures

TABLE OF CONTENTS (Contd)

TABLE OF CONTENTS (Contd)

List of Tables

LIST OF SYMBOLS

All vectors are denoted by lower case letters.
All matrices are denoted by upper case letters.

Symbol	Chap.	Definition
a		q vector of unknown parameters
a	7C	scalar design acceleration
a'		q' vector of unknown parameters except initial conditions
A	2A,B	n × n matrix in performance index
A	2C	scalar independent of x
A		q × q a priori covariance matrix of a.
A	7B	independent variable in one dimensional search
B		p × p matrix in performance index
B_i		m × m matrix, (4.24)
c_1, c_2	8	scalar constants (p. 141)
c_1-c_{12}	9	scalar constants (p. 162)
C	2	p × n feedback gain matrix
C	2E	scalar cost function
C		m × m correlation matrix
D	5	n × m matrix, (5.3)
D		q × q weighting matrix
D	9	scalar constant (p. 162)
e_1, e_2		scalar constants
f		n vector function
F		n × n state matrix

C

LIST OF SYMBOLS (Cont)

All vectors are denoted by lower case letters.
All matrices are denoted by upper case letters.

Symbol	Chap.	Definition
g		N vector
G		$n \times p$ input matrix
G	2C	$n \times n$ process input matrix
h		m vector function
H		$m \times n$ output matrix
H	7B	$N \times N$ conjugate gradient matrix
$\mathcal{H}$		Hamiltonian, (2.4)
I, I_a, I_x		information matrix
I_y, I_{xx}, I_{zz}, I_{xz}		moments and product of inertia
J		performance index
J°		optimal return function
k_{32}, k_{34}, k_{54}		scalar constants
K		$n \times m$ Kalman filter gain matrix
K	5C	scalar gain in steepest descent algorithm
ℓ_β, ℓ_r, ℓ_p, ℓ_{δ_a}		partial derivatives of roll moment with respect to β, r, p, δ_a .
L	4	scalar likelihood ratio
L	5	$m \times m$ matrix, (5.3)
$\mathcal{L}(\cdot)$		operator (2.39)

LIST OF SYMBOLS (Cont)

All vectors are denoted by lower case letters.
All matrices are denoted by upper case letters.

Symbol	Chap.	Definition
m		number of outputs
m	7	magnitude of state constraint
m	8,9	aircraft mass
M		n X n covariance matrix
$M_{\alpha}, M_{\dot{\alpha}}, m_q, M_{\delta e}$		partial derivatives of pitch moment with respect to α, $\dot{\alpha}$, q, δ_e .
n		order of system
$n_{\beta}, n_r, n_p, n_{\delta r}$		partial derivatives of yaw moment with respect to β, r, p, δ_r .
N		number of switch times
p		number of inputs
p(·)		probability distribution of (·) .
p	9	roll angular velocity
P		n X n covariance matrix of x.
P_a		q X q covariance matrix of a.
P	9	phase
q		number of unknown parameters
q		scalar intensity of white noise
q	8	pitch angular velocity
q'		number of unknown parameters, except initial conditions
Q		n X n intensity (or covariance) matrix of w

LIST OF SYMBOLS (Cont)

All vectors are denoted by lower case letters.
All matrices are denoted by upper case letters.

Symbol	Chap.	Definition
r		order of minimal annihilation polynomial
r		scalar density of white noise
r	7	N direction vector
r	9	yaw angular velocity
R		m $\times$ m intensity (or covariance) matrix of v
s		Laplace variable
s	2	scalar function (2.32)
S		n $\times$ n matrix, (2.12)
S_i	7	scalar switching function
t		time
T		length of test
u		p input vector
u_o		forward velocity
v		m white gaussian process (or sequence) vector
w		n white gaussian process (or sequence) vector
x		n state vector
X	2D	n $\times$ n matrix (2.70)
X		set of x
X	7	n $\times$ n covariance matrix of x
y		n + q' augmented state vector
y	3	m output vector
y	7	augmented state vector consisting of state, sensitivity functions, and information matrix

LIST OF SYMBOLS (Cont)

Y_β		partial derivative of lateral force with respect to β
z		m output (measurement) vector
Z		set of z

Greek Symbols

α		q vector of unknown parameters
α	7E	process noise parameter (p. 119)
α	7F	nondimensional number
α	8	angle of attack
β	7E	magnitude parameter (p. 119)
β		sideslip angle
Γ	2C	n $\times$ n process input matrix
Γ		q' adjoint vector
$\delta(\cdot)$		perturbation of $(\cdot)$
$\delta_e, \delta_r, \delta_a$		elevator, rudder, and aileron deflections
Δ		time interval
η		process noise parameter (7.95)
λ		n adjoint vector
λ	3	eigenvalue of F matrix
Λ		n $\times$ n matrix (2.70)
ν		n innovations process vector
σ		standard deviation
τ		nondimensional time
ϕ		performance index

LIST OF SYMBOLS (Cont)

ϕ		n X n state transition matrix
ψ		n + q' augmented adjoint vector
ψ	9	yaw angle
ω		angular frequency

Subscripts

A	approximate
c	control constraint
f	final
i	ith component of a vector
i	value at ith stage
i,j	i,jth component of a matrix
max	maximum value
N	nominal
o	initial

Miscellaneous

$\hat{(\cdot)}$	estimated value or expected value of $(\cdot)$.
$\tilde{(\cdot)}$	error in $(\cdot)$

Chapter I

INTRODUCTION

A. BACKGROUND

This thesis is concerned with determining inputs to identify parameters of a system with the greatest possible accuracy. The theory developed is applied to determining the optimal inputs (elevator, rudder, and aileron deflections vs. time) for an aircraft flight test performed to identify the dynamic stability derivatives of that aircraft. When we consider that flight tests for a large commercial jet aircraft run as high as $50,000 per hour [KR-1], then we can appreciate the importance of designing meaningful flight tests.

There are many approaches to the problem of identifying system parameters from input-output measurements.* Here, we consider systems that can be adequately described by a set of linear differential equations with constant coefficients of the form

$$\begin{aligned} \dot{x} &= Fx + Gu + w \\ z &= Hx + v \end{aligned} \tag{1.1}$$

where x is an n-dimensional state vector, u is a p-dimensional input vector, z is an m-dimensional output vector, w is an n-dimensional white gaussian process with zero mean and intensity matrix Q, and v is an m-dimensional white gaussian measurement process with zero mean and

---

* See, for example, the recent survey paper by Astrom & Eykhoff [AS-1].

intensity matrix R.

In Chapter II we present a brief review of the major results of optimal control and estimation theory which is used in developing the results of this thesis. Estimating parameters in the F, G, H, Q, and R matrices is known as identification and may be viewed as a problem in nonlinear estimation, and the optimal input for identification may be viewed as a stochastic control problem.

The process of describing a system by a set of equations of the form (1.1) is called mathematical modelling. We divide the process into three tasks:

Task 1: Structure Determination. Determine the order n and the structure of the system. A brief introduction to this problem is presented in Chapter III.

Task 2: Identification. Identify the unknown parameters in the model assumed above, according to an identification criterion. Measurements of the inputs and outputs from a previously run test are used in an identification algorithm. A history of identification techniques as applied to the problems of aircraft may be found in Denery [DE-2]. Identification criteria and algorithms are presented in Chapters IV and V respectively.

Task 3: Testing. Design and generate inputs to the system and measure corresponding outputs. Choosing optimal inputs is the subject of Chapter VI through Chapter IX.

B. INPUT DESIGN

In estimating the state of a linear system, the accuracy is independent of the control input, u. However, in estimating parameters of a linear system (a nonlinear estimation problem), the accuracy is dependent on the control input.

If we attempt to choose an optimal input prior to running any tests, a prior estimate of the unknown parameters is required. If these estimates are poor, another test may be required using a revised optimal input. This is the approach used in this thesis as opposed to the more difficult feedback control approach.

The problem of designing optimal inputs for system identification has received recent treatment by Nahi and Wallis [NA-1], Aoki and Staley [AO-1], and Mehra [ME-3]. They also take the approach of designing an input before the test is run, based upon estimates of the parameters to be identified. All of them suggest maximizing the trace of the information matrix which can be a poor criterion. As a better criterion, I suggest minimizing the trace of the inverse of the information matrix. Nahi and Wallace [NA-1] formulate the problem with an amplitude constraint on the input, as done in this thesis. Aoki and Staley [AO-1] and Mehra [ME-3] consider the case of an integral square constraint on the input.

In practice, the input design for aircraft parameter identification is a balance between (1) a good signal which is large enough relative to instrument noise and vehicle disturbances, and, (2) maintaining the instrumentation and the dynamics of the aircraft within their linear regions. If the linear approximation is not a good one for the data obtained from a flight test, then the input is far from optimal in a practical sense. The only constraint considered in the aircraft problem in this thesis (Chapters VIII and IX) has been a control input amplitude constraint. The next step in the solution would be the addition of state inequality constraints to maintain the states within their linear regions. A simpler solution to meet the linearity requirement would be the use of the solution in this thesis, but with the amplitude constraint lowered to meet the linearity requirement.

C. REVIEW BY CHAPTER

In Chapter II, a review of optimal control and estimation theory is presented. A contribution presented in this Chapter is the section on calculating the information matrix for a nonlinear system. The information matrix (whose inverse is a lower bound for the covariance) may be calculated when the covariance itself may not be determined (such as when the initial covariance is large in relation to the nonlinearities).

In Chapter III, some considerations on constructing canonical forms are presented. A comparison is made between Denery's [DE-2] and Spain's [SP-1] canonical forms, with respect to the number of parameters in each form.

In Chapter IV, the maximum a posteriori criterion is developed for the identification problem with noisy measurements of the output. With the addition of process noise, the state equations are replaced by the Kalman filter equations.

In Chapter V, two promising identification algorithms are presented. The first method is Denery's combined algorithm, and the second is a first order gradient algorithm. Both are applied to minimizing the performance indices of Chapter V.

In Chapter VI, we form the information matrix for the unknown parameters to be identified. The input criterion used is the trace of the inverse of the information matrix. A simple example illustrates the fact that maximizing the trace of the information matrix can yield poor results. The information matrix as an input criterion is also developed from the two identification algorithms of the previous Chapter. An interpretation of the sensitivity functions for parameters in F and G is derived from the extended Kalman filter. The Chapter concludes with calculating the information matrix for the case with process noise.

In Chapter VII, we look at optimizing the input criterion developed in Chapter VI. To minimize the trace of the inverse of the information matrix with inequality constraints on the input yields "bang-bang" inputs as optimal. The conjugate gradient algorithm is then used to optimize the criterion with respect to the switch times. For long tests of stable systems, the optimal input may be approximated as a sine wave. The last six sections present six examples: • The first problem is to find the optimal rocket sled acceleration to identify two parameters of an accelerometer. • The next problem is to find the optimal input to identify one parameter of a first order system. • In the next two examples, the first order system is repeated with process noise and with a state inequality constraint. • The last two problems illustrate the nature of optimal inputs for the identification of parameters in unstable systems.

In Chapter VIII, we find the "optimal" elevator input to identify $M_{\dot{\alpha}}$ and M_q of the short period longitudinal dynamics of an aircraft. The switch times and the performance index are plotted as functions of the length of the test. The two unknown parameters are identified using Denery's algorithm from simulated data using optimal and nonoptimal inputs. The simulation verifies the improved performance expected from the optimal input.

In Chapter IX, we find the "optimal" aileron and rudder inputs to identify the four dynamic stability derivatives (ℓ_r, ℓ_p, n_r, n_p) of the lateral equations of motion of an airplane. The only constraint considered was an amplitude constraint on the input. Without the addition of state-inequality constraints, these results must be considered preliminary for all but the shortest of flight tests.

In Chapter X, we present conclusions and recommendations for further research.

Chapter II

REVIEW OF OPTIMAL CONTROL THEORY

A. DETERMINISTIC CONTROL*

In deterministic optimal control theory, a performance index

$$J = \phi[x(t_f)] + \int_{t_0}^{t_f} L(x,u,t)dt \tag{2.1}$$

is minimized by choice of $u(t)$ subject to the constraint

$$\dot{x} = f(x,u,t) \qquad x(t_o) = x_o \tag{2.2}$$

where x is an n-dimensional state vector, and u is a p-dimensional control vector. The calculus-of-variations approach to finding the optimum $u(t)$ yields a two-point-boundary-value problem (TPBVP) specified by (2.2) and the adjoint equation

$$\begin{aligned} \dot{\lambda} &= -\left[\frac{\partial \mathcal{H}}{\partial x}\right]^T \\ \lambda(t_f) &= \left[\frac{\partial \phi}{\partial x(t_f)}\right]^T \end{aligned} \tag{2.3}$$

where the Hamiltonian is defined by

$$\mathcal{H} \triangleq L(x,u,t) + \lambda^T f(x,u,t) \tag{2.4}$$

and the control u is chosen to minimize the Hamiltonian.

---

* The first two sections are based upon Bryson & Ho [BRY-1].

For the special case where the cost function is quadratic in the state and control variables, and the state equations are linear in the state and control variables, we have

$$J = \frac{1}{2}x^T(t_f)S_f x(t_f) + \int_{t_o}^{t_f} \left[\frac{1}{2}x^T Ax + \frac{1}{2}u^T Bu\right] dt \tag{2.5}$$

and

$$x = Fx + Gu, \qquad x(t_o) = x_o \tag{2.6}$$

where A and B are symmetric, A is positive semi-definite and B is positive definite. The Hamiltonian becomes

$$\mathcal{H} = \frac{1}{2} x^T Ax + \frac{1}{2} u^T Bu + \lambda^T(Fx + Gu) \tag{2.7}$$

so that the optimizing control vector is

$$u = -B^{-1}G^T\lambda . \tag{2.8}$$

The two-point-boundary-value problem becomes

$$\begin{aligned} \dot{x} &= Fx - GB^{-1}G^T\lambda, \qquad x(t_o) = x_o \\ \dot{\lambda} &= -Ax - F^T\lambda, \qquad \lambda(t_f) = S_f x(f_f) . \end{aligned} \tag{2.9}$$

This may be solved by the backward sweep method by letting

$$\lambda = Sx \tag{2.10}$$

so that

$$u = -B^{-1}G^T Sx = -Cx . \tag{2.11}$$

S is determined by a matrix Riccati equation

$$\dot{S} = -SF - F^T S - A + SGB^{-1}G^T S, \qquad S(t_f) = S_f . \tag{2.12}$$

The same result may be obtained by dynamic programming where we must solve the Hamilton-Jacobi-Bellman partial differential equation

$$-\frac{\partial J^o}{\partial t} = \min_u \mathcal{H}\left(x, \frac{\partial J^o}{\partial x}, u, t\right)$$

$$J[x(t_f); t_f)] = \phi[x(t_f)] \tag{2.13}$$

for the optimal return function (the performance index expressed as a function of the state x and time t). For the linear-quadratic problem the optimal return function is given by

$$J^o(x,t) = \frac{1}{2} x^T S(t) x . \tag{2.14}$$

B. LINEAR STOCHASTIC CONTROL

For a linear system with state x that is initially $N(x_o, P_o)$ (i.e., gaussian with mean x_o and covariance matrix P_o), driven by white gaussian noise w with zero mean and intensity matrix $Q(t)$ and described by

$$\dot{x} = Fx + Gu + w , \tag{2.15}$$

with measurements z that are corrupted by white gaussian noise v with zero mean and intensity matrix $R(t)$ according to

$$z = Hx + v , \tag{2.16}$$

the conditional probability distribution of the state at time t, given measurements $Z(t_f) = \{z(t),\ t_o \leq t \leq t_f\}$ is gaussian with mean $\hat{x}(t|t_f)$ and covariance $P(t|t_f)$.

For $t \leq t_f$, $\hat{x}(t|t_f)$ and $P(t|t_f)$ are found by minimizing

$$J = \frac{1}{2}[\hat{x}(t_o) - x_o]^T P_o^{-1} [\hat{x}(t_o) - x_o] + \frac{1}{2} \int_{t_o}^{t_f} [w^T Q^{-1} w + (z - Hx)^T R^{-1} (z - Hx)]\, dt \tag{2.17}$$

subject to (2.15) above. This results in the two-point-boundary-value problem

$$\begin{pmatrix} \dot{\hat{x}}(t|t_f) \\ \dot{\lambda} \end{pmatrix} = \begin{bmatrix} F & -Q \\ -H^T R^{-1} H & -F^T \end{bmatrix} \begin{pmatrix} \hat{x}(t|t_f) \\ \lambda \end{pmatrix} + \begin{pmatrix} 0 \\ H^T R^{-1} z \end{pmatrix} \tag{2.18}$$

$$\hat{x}(t_o|t_f) = x_o - P_o \lambda(t_o), \qquad \lambda(t_f) = 0 .$$

This may be solved using the sweep method by letting

$$\hat{x}(t|t_f) = \hat{x} - P\lambda(t) \tag{2.19}$$

where the filtered estimates $\hat{x} \triangleq \hat{x}(t|t)$ and $P \triangleq P(t|t)$ are given by the Kalman-Bucy filter equations

$$\begin{aligned} \dot{\hat{x}} &= F\hat{x} + Gu + PH^T R^{-1}(z - H\hat{x}), \qquad & \hat{x}(t_o) &= x_o \\ \dot{P} &= FP + PF^T + Q - PH^T R^{-1} HP, \qquad & P(t_o) &= P_o \end{aligned} \tag{2.20}$$

and λ is given by

$$\dot{\lambda} = -(F - PH^T R^{-1} H)^T \lambda + H^T R^{-1}(z - H\hat{x}), \qquad \lambda(t_f) = 0 . \tag{2.21}$$

$x(t|t_f)$ is then given by (2.19), and $P(t|t_f)$ is given by

$$P(t|t_f) = P + P\Lambda P \tag{2.22}$$

where Λ is determined by

$$\dot{\Lambda} = -(F - PH^TR^{-1}H)^T\Lambda - \Lambda(F - PH^TR^{-1}H) + H^TR^{-1}H, \quad \Lambda(t_f) = 0 \ . \tag{2.23}$$

For the prediction case where $t > t_1$, $x(t|t_1)$ and $P(t|t_1)$ are determined by

$$\begin{aligned} \dot{\hat{x}}(t|t_1) &= F\hat{x}(t|t_1) + Gu \\ \hat{x}(t_1|t_1) &= \hat{x}(t_1) \ ; \\ \dot{P}(t|t_1) &= FP(t|t_1) + P(t|t_1)F^T + Q \\ P(t_1|t_1) &= P(t_1) \ . \end{aligned} \tag{2.24}$$

If we let our performance index be the ensemble average of a quadratic cost function

$$J = EC = E\left\{\tfrac{1}{2}x^T(t_f)S_fx(t_f) + \int_{t_o}^{t_f} [\tfrac{1}{2}x^TAx + \tfrac{1}{2}u^TBu]dt\right\} \tag{2.25}$$

then the separation theorem tells us that the optimal control is the Kalman-Bucy filter followed by the optimal deterministic feedback controller.

For the optimal control u^o to be realizable, it must be a functional of $Z(t)$ [the measurements up to time t, $z(\tau)$, $t_o \le \tau \le t$], and our initial information about the system. However, this would appear to imply that (2.15) is no longer Markovian and Dynamic Programming techniques (as well as calculus of variations techniques) are no longer applicable [WO-1, p. 211].

We know that $\hat{x}$ and P are sufficient statistics for the stochastic process (2.15) given u; let us assume for the moment that just $\hat{x}$ represents a sufficient statistic to mechanize u^*. We can then define the stochastic optimal return function; expressed as a function of $\hat{x}$ and t, $J^o(\hat{x},t)$ as the minimum of

$$J(\hat{x},t,u) = E\left\{\tfrac{1}{2}x^T(t_f)S_f x(t_f) + \int_t^{t_f} [\tfrac{1}{2}x^T Ax + \tfrac{1}{2}u^T Bu]dt \,\middle|\, Z(t)\right\} \quad (2.26)$$

where $E\{\cdot|Z(t)\}$ represents the ensemble average for that subset of the ensemble with measurements $Z(t)$. Note that the return function (2.26) evaluated at t_o equals the performance index defined in (2.25). Since the "innovations" ν in the Kalman Filter representation

$$\dot{\hat{x}} = F\hat{x} + Gu + K\nu \quad (2.27)$$

is white with intensity R, the stochastic Hamilton-Jacobi-Bellman equation for J^o is

$$\min_u\left\{J_t^o+\tfrac{1}{2}Tr(J^o_{\hat{x}\hat{x}}PH^T R^{-1}HP) + \tfrac{1}{2}\hat{x}^T A\hat{x} + \tfrac{1}{2}TrAP + \tfrac{1}{2}u^T Bu + J_{\hat{x}}(F\hat{x}+Gu)\right\} = 0 \quad (2.28)$$

which becomes

$$J_t^o + \tfrac{1}{2}Tr(J^o_{\hat{x}\hat{x}}PH^T R^{-1}HP) + \tfrac{1}{2}\hat{x}^T A\hat{x} + \tfrac{1}{2}TrAP + J^o_{\hat{x}}F\hat{x} - \tfrac{1}{2}J^o_{\hat{x}}GB^{-1}G^T J^{oT}_{\hat{x}} = 0 \quad (2.29)$$

with the terminal boundary condition

$$J^o[\hat{x}(t_f),t_f] = \tfrac{1}{2}\hat{x}^T(t_f)S_f\hat{x}(t_f) + \tfrac{1}{2}TrS_f P(t_f) \ . \quad (2.30)$$

This has the solution

$$J^o(\hat{x},t) = \tfrac{1}{2}\hat{x}^T S(t)\hat{x} + s(t), \quad (2.31)$$

* After solving the problem with this constraint, Wonham [WO-2] is able to show that the resulting solution is the optimal solution for the unconstrained case.

where S is determined by (2.12) and s is determined by

$$\dot{s} + \tfrac{1}{2}\mathrm{Tr}SPH^TR^{-1}HP + \tfrac{1}{2}\mathrm{Tr}AP = 0$$
$$s(t_f) = \tfrac{1}{2}\mathrm{Tr}S_fP(t_f)\ . \tag{2.32}$$

The optimal control is then $u = -B^{-1}G^TS\hat{x} = -C\hat{x}$ as stated by the separation theorem. The average value of the cost is then

$$J^o[\hat{x}(t_o),t_o] = \tfrac{1}{2}\hat{x}_o^TS(t_o)\hat{x}_o + \tfrac{1}{2}\mathrm{Tr}S_fP(t_f) + $$
$$+ \tfrac{1}{2}\mathrm{Tr}\int_{t_o}^{t_f} SPH^TR^{-1}HP + AP dt. \tag{2.33}$$

By adding the differential $\tfrac{1}{2}$ dSP/dt inside the integral and adding $\tfrac{1}{2}[S(t_o)P(t_o) - S_fP(t_f)]$ outside the integral, we obtain

$$J = \tfrac{1}{2}\mathrm{Tr}\{S(t_o)\hat{X}(t_o) + S(t_o)P(t_o) + $$
$$+ \int_{t_o}^{t_f} SPH^TR^{-1}HP + AP + \dot{S}P + S\dot{P}\ dt\ . \tag{2.34}$$

Substituting into the above equation for $\dot{S}$ and $\dot{P}$, we obtain

$$J = \tfrac{1}{2}\mathrm{Tr}\{S(t_o)X(t_o) + \int_{t_o}^{t_f} SQ + C^TBCP\ dt\}\ . \tag{2.35}$$

C. NONLINEAR ESTIMATION*

For a nonlinear stochastic system and measurements of the form

* This section is based on Sage and Melsa [SA-1] and Jazwinski [JA-1].

$$\dot{x} = f(x,u,t) + G(x,u,t)w\,, \qquad x(t_o) = x_o$$
$$z = h(x,u,t) + v \tag{2.36}$$

we may define an "extended Kalman Filter" by linearizing about the current estimate of the state:

$$\dot{\hat{x}} = f(\hat{x},u,t) + P\frac{\partial h^T}{\partial \hat{x}} R^{-1}[z-h(\hat{x},u,t)], \qquad \hat{x}(t_o) = x_o$$
$$\dot{P} = \frac{\partial f}{\partial \hat{x}} P + P\frac{\partial f^T}{\partial \hat{x}} + GQG^T - P\frac{\partial h^T}{\partial \hat{x}}R^{-1}\frac{\partial h}{\partial \hat{x}}P, \qquad P(t_o) = P_o \tag{2.37}$$

where
$$\frac{\partial f}{\partial \hat{x}} \triangleq \left.\frac{\partial f}{\partial x}\right|_{x=\hat{x}}$$

and
$$\frac{\partial h}{\partial \hat{x}} \triangleq \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}} .$$

Two other promising methods in nonlinear filtering are, conditional mean estimation and maximum a posteriori (conditional mode) estimation.

C.1 Conditional Mean Estimation

The conditional probability distribution of x given $Z(t)$, is given by Kushner's partial differential equation

$$\frac{\partial p}{\partial t} = \mathcal{L}(p) + (h - \hat{h})^T R^{-1}(z - \hat{h})p \tag{2.38}$$

where the operator $\mathcal{L}$ is defined as

$$\mathcal{L}(\cdot) \triangleq -\mathrm{tr}\left\{\frac{\partial f}{\partial x}(\cdot)\right\} + \frac{1}{2}\,\mathrm{tr}\left\{\frac{\partial}{\partial x}\left[\left(\frac{\partial}{\partial x}\right)^T GQG^T(\cdot)\right]\right\}. \tag{2.39}$$

For $R^{-1} = 0$, this reduces to Kolmogorov's partial differential equation which gives the predicted probability distribution in the absence of measurements. Even though there is no known method for solving Kushner's stochastic partial differential equation, it is useful in studying and developing approximate solutions. Also, there is no known expression for the conditional probability distribution of $x(t)$ given later measurements for the nonlinear system given by (2.36). From (2.38), we find that the conditional expectation of a scalar function of x is given by

$$\dot{\hat{\phi}} = \widehat{\phi_x f} + \tfrac{1}{2}\mathrm{tr}\,\widehat{GQG^T\phi_{xx}} + (\widehat{\phi h} - \hat{\phi}\hat{h})^T R^{-1}(z - \hat{h}) \tag{2.40}$$

where the expectation operator $\hat{\cdot}$ is defined by

$$(\hat{\cdot}) \triangleq \int (\hat{\cdot})\, p[x|Z(t)]dx\ . \tag{2.41}$$

From (2.40) we find that the conditional mean and covariance of x are given by

$$\dot{\hat{x}} = \hat{f} + \widehat{(x - \hat{x})h^T} R^{-1}(z - \hat{h}) \tag{2.42}$$

$$\begin{aligned}\dot{P} = {} & \widehat{f(x - \hat{x})} + \widehat{(x - \hat{x})f^T} + \widehat{GQG^T} - \frac{d\hat{x}}{dt}\hat{x}^T \\ & + \widehat{(x - \hat{x})(x - \hat{x})^T(h - \hat{h})R^{-1}(z - \hat{h})}\ .\end{aligned} \tag{2.43}$$

To evaluate (2.42) and (2.43) for the first and second moments of $p(x|Z)$, we would have to know all the moments. An approximate solution for $\hat{x}$ and P may be obtained by expanding $f(x,u,t)$, $h(x,u,t)$, and $G(x,u,t)\ Q(t)\ G^T(x,u,t)$ in a Taylor series. By expanding to second order and using the fact that for nearly gaussian densities

$$E\{\tilde{x}_k\tilde{x}_\ell\tilde{x}_i\tilde{x}_j\} = P_{k\ell}P_{ij} + P_{ik}P_{\ell j} + P_{kj}P_{\ell i} \tag{2.44}$$

we obtain the second order filter

$$\dot{\hat{x}} = f + \frac{1}{2}\frac{\partial^2 f}{\partial \hat{x}^2} : P + P\frac{\partial h^T}{\partial \hat{x}} R^{-1}\left(z - h - \frac{1}{2}\frac{\partial^2 h}{\partial \hat{x}^2} : P\right)$$

$$\dot{P} = \frac{\partial f}{\partial \hat{x}} P + P\frac{\partial f^T}{\partial \hat{x}} - P\frac{\partial h^T}{\partial \hat{x}} R^{-1}\frac{\partial h}{\partial \hat{x}} P + GQG^T + \frac{1}{2}\frac{\partial^2 GQG^T}{\partial \hat{x}^2} : P + \sum \tag{2.45}$$

where

$$\sum_{k\ell} = \frac{1}{2}\sum_{i,j=1}^{N}\left[(P_{ik}P_{\ell j} + P_{kj}P_{\ell i})\frac{\partial^2 h}{\partial \hat{x}_i \partial \hat{x}_j}\right]^T R^{-1}\left(z - h - \frac{\partial^2 h}{\partial \hat{x}^2} : P\right) \tag{2.46}$$

and the operation $:$ is defined by

$$\left[\frac{\partial^2\{\cdot\}}{\partial x^2} : P\right]_{ij} = tr\left[\frac{\partial^2\{\cdot\}_{ij}}{\partial x^2} P\right]. \tag{2.47}$$

C.2 Maximum A Posteriori Estimation

A criterion for the maximum a posteriori estimate of the trajectory of x is obtained for the discrete case and its corresponding continuous criterion is found by a heuristic limiting process. The equivalent discrete system is specified by

$$\begin{aligned} x(k+1) &= \phi[x(k),u(k),k] + \Gamma[x(k),u(k),k]w(k) \\ z(k) &= h[x(k),\ u(k),\ k] + v(k) \end{aligned} \tag{2.48}$$

where $w(k)$ and $v(k)$ are gaussian and

$$\begin{aligned} Ew(k)\ w^T(\ell) &= Q(k)\ \delta_{k\ell} \\ Ev(k)\ v^T(\ell) &= R(k)\ \delta_{k\ell}\ . \end{aligned} \tag{2.49}$$

Let $X(k_f)$ and $Z(k_f)$ denote $x(k_o),\ldots x(k_f)$ and $z(k_1),\ z(k_2),\ldots z(k_f)$ respectively. According to Bayes' rule

$$p[X|Z] = \frac{p[Z|X]p[X]}{p[Z]} . \tag{2.50}$$

Since $v(k)$ is gaussian

$$p[Z|X] = \prod_{k=k_o+1}^{k_f} \frac{1}{\sqrt{(2\pi)^m |R|}} \exp\left\{- \frac{1}{2}(z(k)-h)^T R^{-1}(k)(z(k)-h)\right\}. \tag{2.51}$$

Since $w(k)$ is a white Gauss-Markov sequence

$$p[X] = p[x(k_o)] \prod_{k=k_o+1}^{k_f} p[x(k)|x(k-1)] \tag{2.52}$$

where $p[x(k)|x(k-1)]$ is gaussian with mean $\phi[x(k-1), u(k-1), k-1]$ and covariance

$$\Gamma\left[x(k-1), u(k-1), k-1\right]Q(k-1)\,\Gamma\left[x(k-1), u(k-1), k-1\right] .$$

The conditional probability distribution is then

$$p[X(k_f)|Z(k_f)] = A \exp\left\{-\tfrac{1}{2}\|x(t_o)-x_o\|^2_{P_o^{-1}} - \frac{1}{2}\sum_{k=k_o+1}^{k_f}\left[\|z(k)-h\|^2_{R^{-1}(k)} + \|x(k) - \phi[x(k-1),u(k-1),k-1]\|^2_{(\Gamma Q\Gamma^T)^{-1}}\right]\right\} \tag{2.53}$$

where A is independent of x. Maximizing the conditional probability distribution is equivalent to minimizing the performance index

$$J = \frac{1}{2}\|x(k_o)-x_o\|^2_{P_o^{-1}} + \frac{1}{2}\sum_{k=k_o}^{k_f} \|z(k+1) - h[x(k+1),u(k+1),k+1]\|^2_{R^{-1}(k)} + \|w(k)\|^2_{Q^{-1}(k)} . \tag{2.54}$$

This criterion yields the maximum a posteriori estimate for the joint probability distribution of $x(k_o)$, $x(k_1), \dots x(k_f)$. The value of $x(k)$ found by minimizing (2.54) is not necessarily the mode of the marginal probability distribution for $x(k)$. In principle, we could obtain the marginal probability distribution for $x(k)$ by integrating the joint probability distribution with respect to $x(o), x(1), \dots x(k-1), x(k+1), \dots x(N)$. Passing to the limit, the maximum a posteriori criterion for the continuous system for the trajectory $X(t_f) \triangleq \{x(\tau),\ t_o \leq \tau \leq t_f\}$ is

$$J = \frac{1}{2}[x(t_o) - x_o]^T P_o^{-1} [x(t_o) - x_o] + \frac{1}{2}\int_{t_o}^{t_f}\left\{[z - h]^T R^{-1}[z - h] + w^T Q^{-1} w\right\}dt . \tag{2.55}$$

A calculus-of-variations solution leads to the two-point-boundary-value problem

$$\begin{aligned} \dot{x} &= f(x,u,t) - G(x,u,t)Q(t)G(x,u,t)\lambda, & x(t_o) &= x_o - P_o\lambda(t_o) \\ \dot{\lambda} &= -\left[\frac{\partial f}{\partial x}\right]^T\lambda + \left[\frac{\partial h}{\partial x}\right]^T R^{-1}[z - h], & \lambda(t_f) &= 0 . \end{aligned} \tag{2.56}$$

An approximate solution to this two-point-boundary-value problem can be solved by means of invariant imbedding leading to

$$\begin{aligned} \dot{\hat{x}} &= f(\hat{x},u,t) + P\frac{\partial h^T}{\partial \hat{x}} R^{-1}[z - h(\hat{x},u,t)], & \hat{x}(t_o) &= x_o \\ \dot{P} &= \frac{\partial f}{\partial \hat{x}} P + P\left[\frac{\partial f}{\partial \hat{x}}\right]^T + P\left[\frac{\partial}{\partial \hat{x}}\left\{\frac{\partial h^T}{\partial \hat{x}} R^{-1}(z-h)\right\}\right]P + GQG^T, & P(t_o) &= P_o . \end{aligned} \tag{2.57}$$

Approximate smoothing algorithms can also be obtained in a fashion similar to the filter algorithms. These require the results of the approximate filter solutions.

D. AN INFORMATION MATRIX APPROACH

The approximate filters of the previous section were derived on the assumption that the covariance is "small" compared to the nonlinearities in f, G, and h. For example, in the scalar case, a "smallness" criterion could be obtained by expanding $f(x)$ to second order about $\bar{x}$:

$$f(x) = f(\bar{x}) + f_x\Big|_{x=\bar{x}} (x-\bar{x}) + \frac{1}{2} f_{xx}\Big|_{x=\bar{x}} (x-\bar{x})^2 + \cdots .$$

If the range of $x-\bar{x}$ were $\pm 3\sigma$, then we would have to satisfy the condition

$$\sigma_x^2 \ll \frac{4f_x^2}{9f_{xx}^2} .$$

for the variance of x to be "small." Similar conditions would have to hold for higher order terms in the Taylor series.

If the initial covariance did not meet this smallness requirement, we could still solve (2.56) by some other technique. However, we would still not have an estimate of the covariance of the state. Such an estimate may be obtained by calculating the information matrix.

The Fisher information matrix corresponding to a probability distribution $p(x)$ is defined as follows: [VA-1, Part 1]

$$I_x \triangleq -E\,\frac{\partial^2 \ell n p(x)}{\partial x^2} \tag{2.58}$$

where the expectation operator is defined as $E(\cdot) \triangleq \int_{-\infty}^{+\infty} (\cdot)p(x)dx$.

If x has a gaussian distribution with mean $\bar{x}$ and covariance P, then

$$p(x) = \frac{1}{\sqrt{(2\pi)^n |P|}} \exp\left\{-\tfrac{1}{2}(x - \bar{x})^T P^{-1}(x - \bar{x})\right\} \tag{2.59}$$

and the above definition shows us that

$$I = P^{-1} . \quad (2.60)$$

A general performance index of the form

$$J = \phi_f[x(t_f),t_f] + \phi_o[x(t_o),t_o] + \int_{t_o}^{t_f} L(x,u,t)dt \quad (2.61)$$

may be written as

$$J = J^{(-)} + J^{(+)} \quad (2.62)$$

where $J^{(-)}$ and $J^{(+)}$ are defined by

$$J^{(-)}(t) = \phi_o[x(t_o),t_o] + \int_{t_o}^{t} L(x,u,t)dt$$

$$J^{(+)}(t) = \phi_f[x(t_f),t_f] + \int_{t}^{t_f} L(x,u,t)dt . \quad (2.63)$$

The adjoint variables are equal to [BR-1]

$$\lambda^T(t) = \frac{\partial J^{(+)}(t)}{\partial x(t)} \text{ and } \lambda^T(t) = -\frac{\partial J^{(-)}(t)}{\partial x(t)} . \quad (2.64)$$

Let us make the assumption that the conditional probability distribution of $x(t)$ given measurements $Z(t_f)$, is given by*

$$p(x) = Ae^{-J(x)} \quad (2.65)$$

where A is independent of x.

---

* This is not strictly true since J is the maximum a posteriori criterion for the trajectory $X(t)$ and not a criterion for the marginal probability distribution, $x(t)$.

ε

The information matrix for $x(t)$, given measurements $Z(t_f)$, may then be expressed as a function of the performance index (2.55) by

$$I_x(t|t_f) = E \frac{\partial^2 J}{\partial x^2(t)}\bigg|_{x=\hat{x}} . \tag{2.66}$$

Since

$$\frac{\partial J}{\partial x(t_f)} = \frac{\partial J^{(-)}(t_f)}{\partial x(t_f)} = -\lambda^T(t_f) , \tag{2.67}$$

we have

$$I_x(t_f) \triangleq I_x(t_f|t_f) = -E \frac{\partial \lambda(t_f)}{\partial x(t_f)} . \tag{2.68}$$

The sensitivity matrix

$$E \frac{\partial \lambda(t)}{\partial x(t_f)}$$

is specified by the linear matrix two-point-boundary-value problem

$$\begin{aligned} \dot{X} &= \left(\left[\frac{\partial f}{\partial x}\right] - M\right)X - GQG^T\Lambda, & X(t_f) &= I \\ \dot{\Lambda} &= \left(-N - \frac{\partial h^T}{\partial x} R^{-1} \frac{\partial h}{\partial x}\right)X - \left[\frac{\partial f}{\partial x}\right]^T\Lambda, & \Lambda(t_o) &= -P_o^{-1}X(t_o) \end{aligned} \tag{2.69}$$

where

$$X(t) \triangleq E\frac{\partial x(t)}{\partial x(t_f)} \quad \text{and} \quad \Lambda(t) \triangleq E \frac{\partial \lambda(t)}{\partial \lambda(t_f)} , \tag{2.70}$$

the ith row of $M = \lambda^T(\partial m_i/\partial x)$, where m_i^T = ith row of GQG^T, and the ith row of $N = \lambda^T(\partial n_i/\partial x)$, where n_i^T = ith row of $[(\partial f/\partial x)]^T$. Once the TPBVP of (2.56) is solved, the coefficients in (2.69) may be evaluated.

D.1 Linear System

As an example, consider the linear system and measurements specified by

$$\dot{x} = Fx$$
$$z = Hx + v \tag{2.71}$$

with the performance index

$$J = \tfrac{1}{2}[x(t_o)-x_o]^T P_o^{-1}[x(t_o)-x_o] + \tfrac{1}{2}\int_{t_o}^{t_f} (z-Hx)^T R^{-1}(z-Hx)dt \ . \tag{2.72}$$

The vector TPBVP for x and λ is

$$\begin{aligned} \dot{x} &= Fx & \lambda(t_o) &= -P_o^{-1}[x(t_o) - x_o] \\ \dot{\lambda} &= -F^T\lambda + H R^{-1}(z-Hx), & \lambda(t_f) &= 0 , \end{aligned} \tag{2.73}$$

and the matrix TPBVP is then

$$\begin{aligned} \dot{X} &= FX, & X(t_f) &= I \\ \dot{\Lambda} &= -F^T\Lambda - H^T R^{-1} HX, & \Lambda(t_o) &= -P_o^{-1}X(t_o) \end{aligned} \tag{2.74}$$

and the information matrix is

$$I_x(t_f) = -\Lambda(t_f) \ . \tag{2.75}$$

Now let us verify that this answer agrees with what the Kalman filter would give: Let

$$\begin{bmatrix} \phi_{xx} & \phi_{x\lambda} \\ \phi_{\lambda x} & \phi_{\lambda\lambda} \end{bmatrix}$$

be the transition matrix for

$$\begin{bmatrix} F & 0 \\ -H^T R^{-1} H & -F^T \end{bmatrix}$$

so that

$$X(t_o) = \phi_{xx}(t_o, t_f)X(t_f) \ , \tag{2.76a}$$

$$\Lambda(t_f) = \phi_{\lambda x}(t_f, t_o)\, X(t_o) + \phi_{\lambda\lambda}(t_f, t_o)\, \Lambda(t_o) \; . \tag{2.76b}$$

Making the substitutions

$$X(t_f) = I \quad \text{and} \quad \Lambda(t_o) = -P_o^{-1} X(t_o) \tag{2.77}$$

we have

$$\Lambda(t_f) = [\phi_{\lambda x}(t_f, t_o) - \phi_{\lambda\lambda}(t_f, t_o) P^{-1}]\, \phi_{xx}(t_o, t_f) \; . \tag{2.78}$$

Differentiating

$$\frac{\partial}{\partial t_f} \Lambda(t_f) = \left[-H^T R^{-1} H \phi_{xx}(t_f, t_o) - F^T \phi_{\lambda x}(t_f, t_o) + F^T \phi_{\lambda\lambda}(t_f, t_o) P_o^{-1}\right]$$

$$\times\ \phi_{xx}(t_o, t_f) + \left[\phi_{\lambda x}(t_f, t_o) - \phi_{\lambda\lambda}(t_f, t_o) P_o^{-1}\right] \phi_{xx}(t_o, t_f)(-)F$$

and simplifying

$$\frac{\partial}{\partial t_f} \Lambda(t_f) = -H^T R^{-1} H - F^T \Lambda(t_f) - \Lambda(t_f) F \tag{2.79}$$

we find that I_x satisfies the equation for P^{-1} in the Kalman filter:

$$\dot{I}_x = -I_x F - F^T I_x + H^T R^{-1} H, \qquad I_x(t_o) = P_o^{-1} \; . \tag{2.80}$$

For this simple example, a direct* derivation of J is easier; readily leading to

$$I_x(t_f) = \frac{\partial^2 J}{\partial x^2(t_f)} = X^T(t_o) P^{-1} X(t_o) + \int_{t_o}^{t_f} X^T(t) H^T R^{-1} H X(t)\, dt \tag{2.81}$$

and only the first equation in (2.74) is needed. Differentiating, we have

---

* By "direct" we mean that the performance index is differentiated directly without employing the adjoint variables.

$$\frac{\partial}{\partial t_f} I_x(t_f) = \dot{X}^T(t_o)P_o^{-1} x(t_o) + x^T(t_o)P_o^{-1} \dot{X}(t_o) + X^T(t_f)H^T R^{-1} HX(t_f)$$

$$+ \int_{t_o}^{t_f} \dot{X}^T H^T R^{-1} HX + X^T H^T R^{-1} H\dot{X} dt. \tag{2.82}$$

If we make the substitutions

$$\dot{X}(t) \triangleq \frac{\partial}{\partial t_f}\left(\frac{\partial x(t)}{\partial x(t_f)}\right) = -\frac{\partial x(t)}{\partial x(t_f)} F = -X(t)F$$

and (2.83)

$$X(t_f) = I$$

we obtain (2.80).

E. NONLINEAR STOCHASTIC CONTROL

If we assume that $\hat{x}$ and P given by (2.37), (2.45), or (2.57) represent a set of sufficient statistics for $p(x,t|Z)$, and $z - h(x,u,t)$ is approximately white with intensity R, then we can form the stochastic Hamilton-Jacobi-Bellman equation. This makes the problem nearly impossible to solve. If we cannot make assumptions such as this there is no known "exact" method of solving the nonlinear stochastic control problem.

The performance index for the nonlinear problem may also include weights upon the moments of the cost as well as just the mean of the cost:

$$J = \alpha_1 EC + \alpha_2 E(C - \hat{C}) + \cdots \alpha_n E(C - \hat{C})^n + \cdots . \tag{2.84}$$

In practice, this performance index could be expanded to second order as is done in the second-order filter.

Chapter III

STRUCTURE DETERMINATION

A. INTRODUCTION

Recall from Chapter I that the first task of mathematical modelling is the determination of the system structure. In many applications, the order and structure of the differential equations may be derived from physical principles. Such is the case in deriving the equations of motion of an airplane.

In more complex systems such as biological or economic processes, the underlying processes are not well known. In such cases, an approximate model of the system may be obtained by assuming a given order or other structural information about the system and fitting data to it.

Let us assume that the structural information about the system may be specified by a set of model numbers. An example of a model number, other than the order of the system n, the number of inputs p, and the number of outputs m, would be the order r of the minimal annihilation polynomial.* A possible method of determining the structure of a system is the following:

(1) Assume a given value for the model numbers (for example, assume a first order system).

(2) Perform the other two tasks of mathematical modelling under this assumption, namely (a) choosing an input and measuring the corresponding output, and (b) identifying the parameters of the assumed structure from input-output records.

---

* A polynomial is an annihilation polynomial if it equals 0 when the F matrix is substituted for the independent variable. The Hamilton-Cayley theorem tells us that the $\underline{n}$th order polynomial of the characteristic equation is an annihilation polynomial. However, there may be other polynomials of lower order that are also annihilation polynomials.

3) Increase the values of the model numbers (for example, increase the order by one) until a structure criterion is met. Two possible structure criteria are: (a) the residuals (difference between the measured output and model output) are "close" to being white. Such a criterion has been used by Mehra [ME-1]; (b) There is no significant reduction in the identification criterion. A significance test for the reduction is given in Astrom and Eykhoff [AS-1]. The latter criterion appears to be the more decisive [SP-1] but requires an identification at one higher value of the model numbers than the former criterion.

The next section discusses useful results from realization theory that may be applied to constructing canonical forms. It also discusses the construction of canonical forms with four model numbers (m, n, p, and r) and compares the canonical forms of Denery and Spain.

B. REALIZATION THEORY*

Realization theory for deterministic systems is concerned with specifying the internal description of a system (i.e., specifying its differential equations) from a known external description of a system (as expressed by its impulse response matrix or transfer function matrix). For the deterministic system

$$\begin{aligned} \dot{x} &= Fx + Gu \\ y &= Hx \end{aligned} \qquad (3.1)$$

with zero initial conditions, the output is given by

$$y(t) = \int_{t_o}^{t} H\phi(t,\tau)G\ u(\tau)d\tau \qquad (3.4)$$

or in the frequency domain, by

$$y(s) = H(sI - F)^{-1}Gu(s)\ . \qquad (3.\ )$$

---

* This section based on Kalman [KAL-1].

As far as any input-output relationships are concerned (with zero initial conditions), the descriptions in (3.4) and (3.5) are equivalent to the description in (3.3). However, the specification of (F, G, H) from either (3.4) or (3.5) is not unique. Before proceeding with the main results of realization theory for linear time invariant systems, two definitions and one theorem are in order.

1. Definition 1:

(F, G, H) is strictly algebraically equivalent to $(\bar{F}, \bar{G}, \bar{H})$ if and only if there exists a non-singular constant matrix T, such that

$$\begin{aligned} \bar{F} &= TFT^{-1} \\ \bar{G} &= TG \\ \bar{H} &= HT^{-1} . \end{aligned} \tag{3.6}$$

2. Definition 2:

(F, G, H) is a minimal realization if there is no other realization $(\bar{F}, \bar{G}, \bar{H})$ with an $\bar{F}$ of order smaller than the order of F.

3. Canonical Structure Theorem

The state vector may be transformed into four mutually exclusive parts (see Fig. 3.1):

Part A: controllable but unobservable;

Part B: controllable and observable;

Part C: uncontrollable and unobservable;

Part D: uncontrollable and observable;

so that F, G, and H take the canonical forms

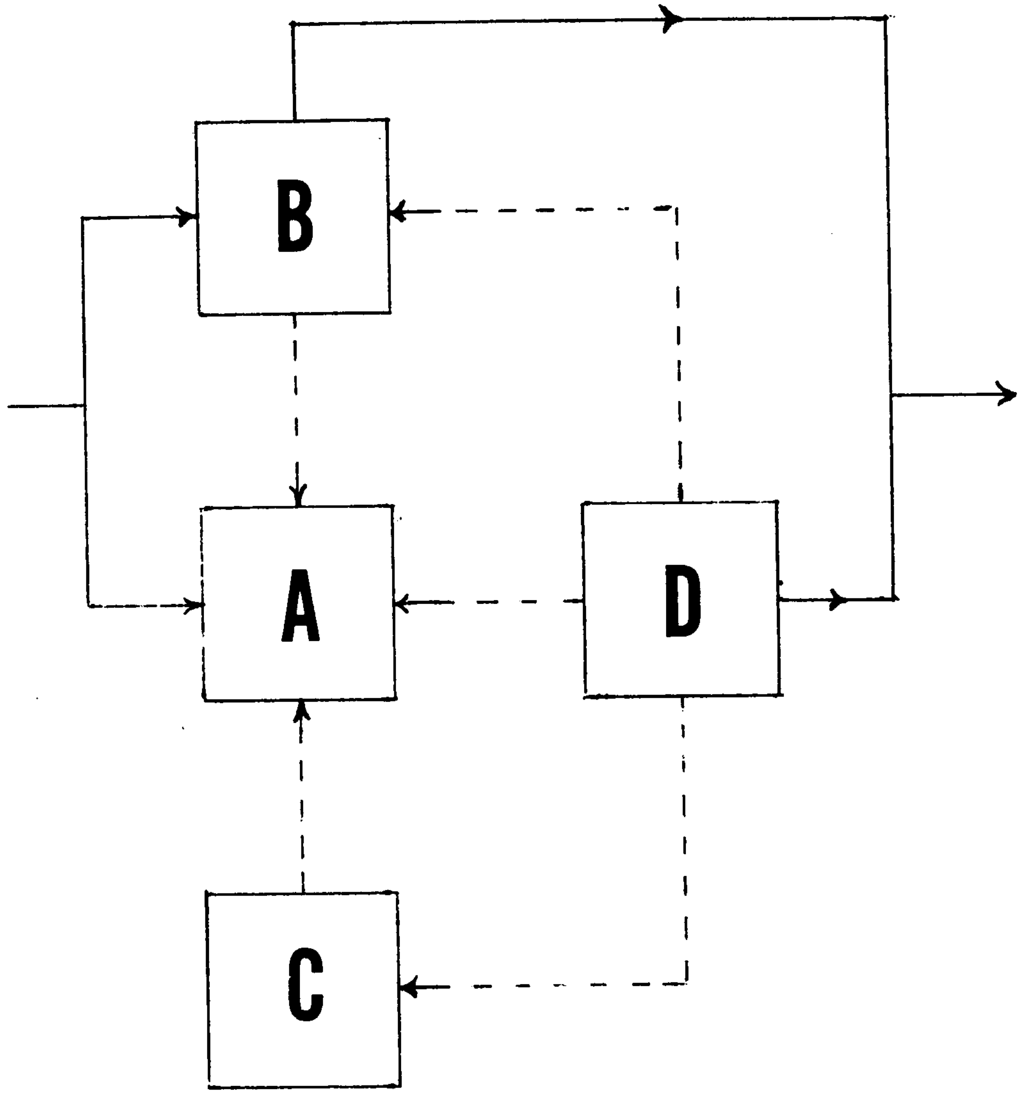


FIG. 3.1. DIAGRAM OF CANONICAL STRUCTURE THEOREM

$$F = \begin{bmatrix} F^{AA} & F^{AB} & F^{AC} & F^{AD} \\ 0 & F^{BB} & 0 & F^{BD} \\ 0 & 0 & F^{CC} & F^{CD} \\ 0 & 0 & 0 & F^{DD} \end{bmatrix};$$

$$G = \begin{bmatrix} G^{A} \\ G^{B} \\ 0 \\ 0 \end{bmatrix}; \qquad (3.7)$$

$$H = \begin{bmatrix} 0 & H^{B} & 0 & H^{D} \end{bmatrix}.$$

From the above theorem it is easy to see that the transfer function for a general system is given by $H^{B}(sI - F^{BB})^{-1}G^{B}$ so that we have:

Result 1: Only the controllable and observable portion of a system can be identified. We must not be too confident that we "know" a system from a description of its input and output. There may be other important parts of the system that we know nothing about.

Conversely, we have,

Result 2: A realization is minimal if and only if it is controllable and observable. We may generate a realization that contains parts A, B, C, and D. However, a minimal realization consists of only part B of the above nonminimal realization.

Finally, we have,

Result 3: Any two minimal realizations (of a time invariant system) are strictly algebraically equivalent. Algorithms for finding a minimal realization are given by Gilbert [GI-1], Kalman [KAL-1], Ho and Kalman [HO-1], and Silverman [SI-1].

There are at least three main criticisms of the realization theory approach to mathematical modelling: (1) The transfer function (or impulse response) matrix has to be determined before it can be applied. Why not identify F, G, and H directly from measurements of the inputs and outputs without first calculating the transfer function matrix? (2) It is assumed that the transfer function (or impulse response) matrix is given exactly; whereas with these external descriptions, the parameters in F, G, and H may be very sensitive to small errors in the transfer function (or impulse response) matrix. (3) One may be led to believe that an impulse or sine input is the "proper" input to use.

C. MINIMAL PARAMETER SET

In parameter identification the number of independent parameters q, needed to describe a system, is of great interest. If a realization is of minimal order, any desired canonical form can be used to enumerate the number of independent parameters. The information matrix provides a means of verifying the identifiability of a set of parameters. The independence of a set of parameters in the information matrix is equivalent to the identifiability of the parameters. If the information matrix for a set of parameters is singular for any input, then we do not have a canonical form.

By knowing the order n, number of inputs p, number of outputs m, and part of the structure of a system, Denery [DE-2] constructs a canonical form involving $n(m + p)$ parameters. The structural information needed consists of the first n linearly independent rows of the observability matrix. If we do not know the first n linearly independent rows, then we must examine each possibility for a given value of n.

For systems with an annihilation polynomial of degree r (but of unknown order $n \geqslant r$), Spain [SP-1] constructs a canonical form involving $r(mp + 1)$ parameters. If F has an annihilating polynomial of degree less than n, then F is similar to a quasidiagonal matrix that has two or more Jordan blocks with the same eigenvalue. It would then seem to be a special case for a physical system to have $r < n$. Thus, Spain's number of parameters is much larger (for multi-input multi-output systems) than Denery's, except for special cases. However, Spain does not assume any structural information and would not have to investigate a large number of cases for each value of r.

Any square matrix with multiple eigenvalues is similar to a quasidiagonal matrix where each diagonal matrix is a Jordan matrix. The possibility of multiple eigenvalues suggests that this form gives us a form with the minimum number of parameters. It is instructive to calculate the number of parameters needed to describe a quasidiagonal canonical form for the model numbers (m, n, p, r). The results are shown in Table 3.1 for $n = 1, 2, 3$. For $n \geq 4$, the number of cases increases greatly; for example, for $n = 4$, there are 14 different cases and for $n = 5$ there are 29 different cases. For each case, the number of parameters is less than or equal to that given by Denery or Spain. (Since each of these cases assumes more about the system.) A method of calculating the results shown in Table 3.1 is illustrated in the following example: Find the number of parameters needed to describe a second order system with two inputs and two outputs. There are three different cases:

Case 1: Distinct eigenvalues. See Fig. 3.2a, $(r = 2)$. As far as input-output relationships are concerned, we could make the following replacements:

$$g_{11} \rightarrow g_{11} h_{11} \qquad h_{11} \rightarrow 1$$

$$g_{12} \rightarrow g_{12} h_{22} \qquad h_{21} \rightarrow h_{21}/h_{11}$$

$$g_{21} \rightarrow g_{21} h_{11} \qquad h_{12} \rightarrow h_{12}/h_{22}$$

$$g_{22} \rightarrow g_{22} h_{22} \qquad h_{22} \rightarrow 1.$$

For this case there are eight parameters: λ_1, λ_2, g_{11}, g_{12}, g_{21}, g_{22}, h_{12}, h_{21} . The information matrix for these eight parameters is nonsingular.

Case 2: Jordan form. See Fig. 3.2b, $(r = 2)$. In this case we make the following replacements:

<u>Table 3.1</u>

The minimal number of parameters, q, of a canonical form for the model numbers (m, n, p, r). Cases are shown for n = 1, 2, 3. For each case, q is shown versus m and p.

Order and Case	F Matrix	q vs. m and p

n = 1, Case 1, r = 1

$$F = \begin{bmatrix} \lambda \end{bmatrix}$$

m \ p	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

n = 2, Case 1, r = 2

$$F = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

m \ p	1	2	3	4
1	4	6	8	10
2	6	8	10	12
3	8	10	12	14
4	10	12	14	16

n = 2, Case 2, r = 2

$$F = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

m \ p	1	2	3	4
1	4	6	8	10
2	6	8	10	12
3	8	10	12	14
4	10	12	14	16

n = 2, Case 3, r = 1

$$F = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

m \ p	1	2	3	4
1				
2		5	7	9
3		7	9	11
4		9	11	13

Contd.

m = number of outputs
n = order of system
p = number of inputs
r = order of minmal annihilation polynomial

Not Observable (hatched cells: row m = 1)

Not Controllable (hatched cells: column p = 1)

F

Table 3.1 (Contd)

<table>
<tr><th>Order and Case</th><th>F Matrix</th><th>q vs. m and p</th></tr>
<tr>
<td>n = 3
Case 1
r = 3
[n = 1, Case 1 ⊕
n = 2, Case 1]</td>
<td>$\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$</td>
<td><table>
<tr><th>m \ p</th><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>1</td><td>6</td><td>9</td><td>12</td><td>15</td></tr>
<tr><td>2</td><td>9</td><td>12</td><td>15</td><td>18</td></tr>
<tr><td>3</td><td>12</td><td>15</td><td>18</td><td>21</td></tr>
<tr><td>4</td><td>15</td><td>18</td><td>21</td><td>24</td></tr>
</table></td>
</tr>
<tr>
<td>n = 3
Case 2
r = 3
[n = 1, Case 1 ⊕
n = 2, Case 2]</td>
<td>$\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_2 \end{bmatrix}$</td>
<td><table>
<tr><th>m \ p</th><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>1</td><td>6</td><td>9</td><td>12</td><td>15</td></tr>
<tr><td>2</td><td>9</td><td>12</td><td>15</td><td>18</td></tr>
<tr><td>3</td><td>12</td><td>15</td><td>18</td><td>21</td></tr>
<tr><td>4</td><td>15</td><td>18</td><td>21</td><td>24</td></tr>
</table></td>
</tr>
<tr>
<td>n = 3
Case 3
r = 3</td>
<td>$\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$</td>
<td><table>
<tr><th>m \ p</th><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>1</td><td>6</td><td>9</td><td>12</td><td>15</td></tr>
<tr><td>2</td><td>9</td><td>12</td><td>15</td><td>18</td></tr>
<tr><td>3</td><td>12</td><td>15</td><td>18</td><td>21</td></tr>
<tr><td>4</td><td>15</td><td>18</td><td>21</td><td>24</td></tr>
</table></td>
</tr>
<tr>
<td>n = 3
Case 4
r = 2
[n = 1, Case 1 ⊕
n = 2, Case 3]</td>
<td>$\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_2 \end{bmatrix}$</td>
<td><table>
<tr><th>m \ p</th><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>1</td><td></td><td></td><td></td><td></td></tr>
<tr><td>2</td><td></td><td>9</td><td>12</td><td>15</td></tr>
<tr><td>3</td><td></td><td>12</td><td>15</td><td>18</td></tr>
<tr><td>4</td><td></td><td>15</td><td>18</td><td>21</td></tr>
</table></td>
</tr>
</table>

Contd.

Table 3.1 (Contd)

<table>
<tr><th>Order and Case</th><th>F Matrix</th><th>q vs. m and p</th></tr>
<tr>
<td>n = 3
Case 5
r = 2</td>
<td>$$\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$$</td>
<td>
<table>
<tr><th>m \ p</th><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>1</td><td></td><td></td><td></td><td></td></tr>
<tr><td>2</td><td></td><td>8</td><td>11</td><td>14</td></tr>
<tr><td>3</td><td></td><td>11</td><td>14</td><td>17</td></tr>
<tr><td>4</td><td></td><td>14</td><td>17</td><td>20</td></tr>
</table>
</td>
</tr>
<tr>
<td>n = 3
Case 6
r = 1</td>
<td>$$\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$</td>
<td>
<table>
<tr><th>m \ p</th><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>1</td><td></td><td></td><td></td><td></td></tr>
<tr><td>2</td><td></td><td></td><td></td><td></td></tr>
<tr><td>3</td><td></td><td></td><td>10</td><td>13</td></tr>
<tr><td>4</td><td></td><td></td><td>13</td><td>16</td></tr>
</table>
</td>
</tr>
</table>

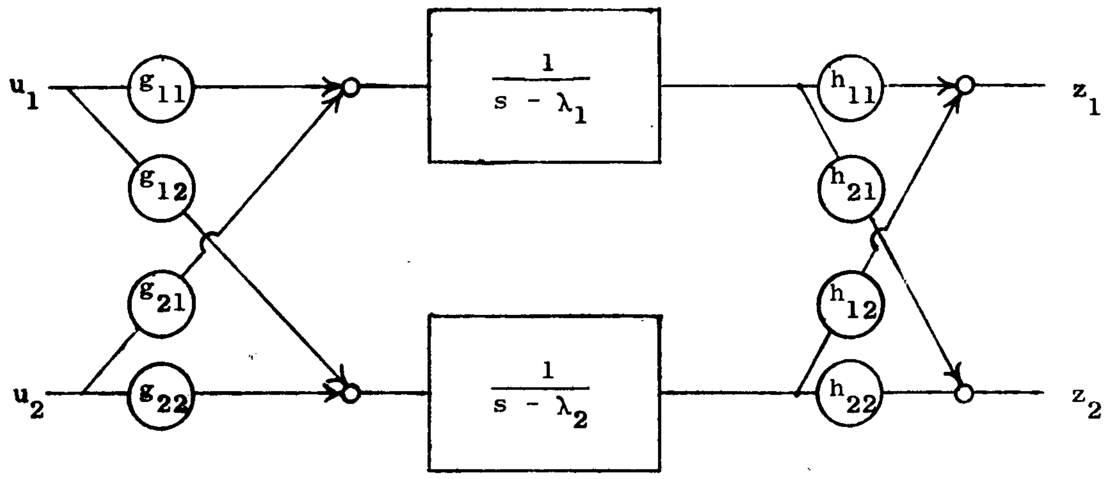


Fig. 3.2a

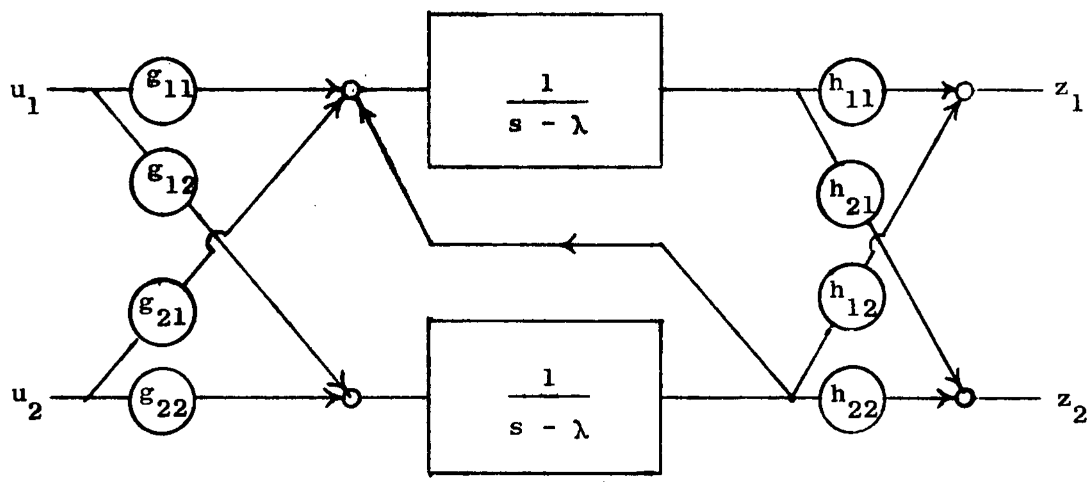


Fig. 3.2b

FIG. 3.2. Schematic diagrams for an example in determining the minimum number of parameters of a canonical form. The example was a second order system with two inputs and two outputs.

$$g_{11} \rightarrow g_{11} h_{11} \qquad h_{11} \rightarrow 1$$

$$g_{12} \rightarrow g_{12} h_{11} \qquad h_{21} \rightarrow h_{21}/h_{11}$$

$$g_{21} \rightarrow g_{21} h_{11} \qquad h_{12} \rightarrow h_{12}/h_{11}$$

$$g_{22} \rightarrow g_{22} h_{11} \qquad h_{22} \rightarrow h_{22}/h_{11} .$$

In this case we cannot normalize with respect to h_{22} due to the extra coupling; however, both eigenvalues are the same so that eight parameters are still all that is necessary; namely, λ, g_{11}, g_{12}, g_{21}, h_{12}, h_{21}, h_{22}.

Case 3: Two (1 x 1) Jordan blocks have the same eigenvalue. See Fig. 3.2c, (r = 1). From the results of Case 1, we know that seven parameters are sufficiently general; but perhaps they are not all identifiable. From Fig. 3.2c, we see that as far as the paths from u_1 to z_1 are concerned, we cannot tell from measurements of the input and output whether we took path $g_{11} \rightarrow 1$ or $g_{12} \rightarrow h_{12}$. We may eliminate one path by setting $h_{12} = 0$ (if it is not needed by some other connection). In going from u_2 to z_2, we reach a similar conclusion about h_{21}. In going from u_1 to z_2, we have to keep either $g_{12} \neq 0$ or $h_{21} \neq 0$; let us choose $g_{12} \neq 0$ and $h_{21} = 0$. From u_2 to z_1, we reach a similar conclusion about setting $h_{12} = 0$. We thus have the possible form shown in Fig. 3.2d, with five parameters: λ, g_{11}, g_{12}, g_{21}, g_{22}. The information matrix for seven parameters can be shown to be singular for any input. This is a consequence of the linear dependence of the sensitivity equations when $\lambda_1 = \lambda_2$. For the set of five parameters, the information matrix is nonsingular.

Although the results in this example were derived assuming that the eigenvalues were real, we would get the same number of parameters since for each complex eigenvalue, its conjugate is also an eigenvalue. Note

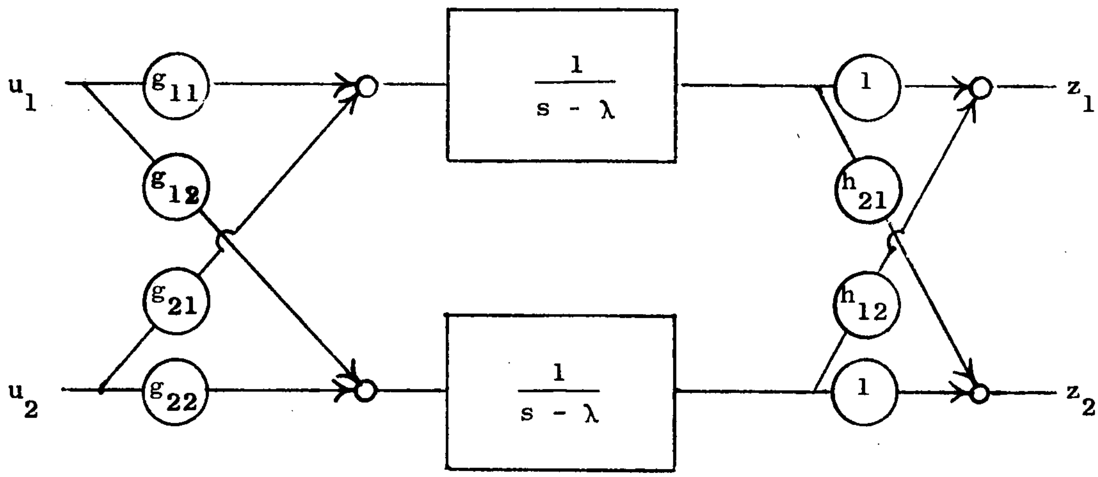


Fig. 3.2c

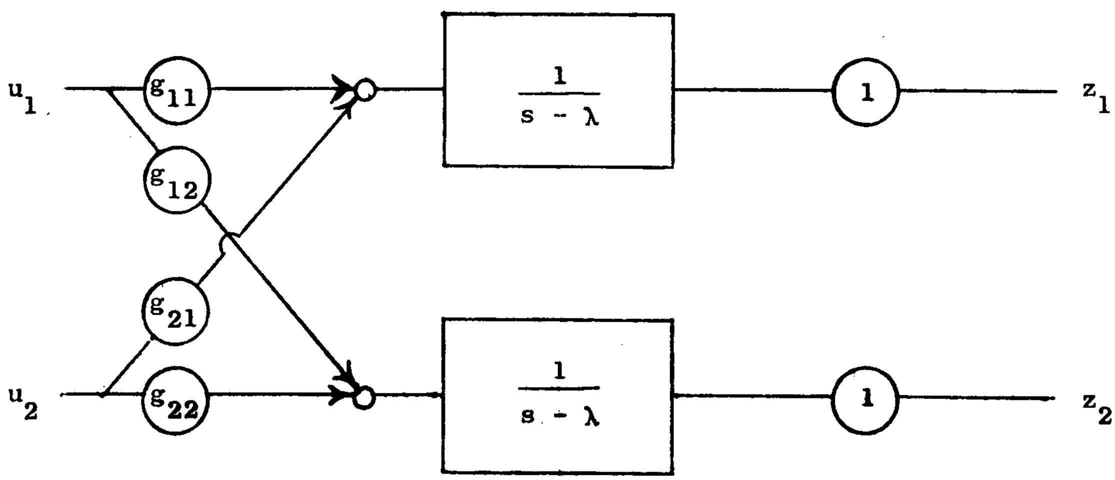


Fig. 3.2d

FIG. 3.2. (Cont) Schematic diagrams for an example in determining the minimum number of parameters of a canonical form. The example was a second order system with two inputs and two outputs.

that for all cases for which no Jordan blocks have the same eigenvalue (i.e., for which $r = n$), the number of parameters is the same as that given by Denery's canonical form, $q = n(m + p)$.

Future research would be useful in determining the best model numbers for multi-input multi-output systems. Considerations should answer the following two questions: (1) What is the minimal number of parameters, q, needed to designate an arbitrary member of the class defined by the model numbers? (2) As the order of the system increases, how many different cases, c, must be examined? In general, the more model numbers we have, the smaller q is but the larger c is. Some optimum trade-off should be possible.

Chapter IV

IDENTIFICATION CRITERIA

A. INTRODUCTION

Let the vector, a, represent the unknown parameters in F, G, H, Q, and R (and the initial conditions), and Z(t) the set of measurements up to time t. The identification criteria developed in this Chapter are based on finding the value of a at the maximum of the a posteriori probability distribution $p_{a|Z}$:

$$a = \arg \max_a p_{a|Z} \ .$$

This is a mathematically simpler approach than the conditional mean approach summarized in Chapt. II.C. Since a is a vector of constant parameters, we do not have the problem noted in Chapt. II.C that there may be a difference between a maximum a posteriori criterion for the joint probability distribution and the marginal probability distribution.

Since Bayes formula tells us that

$$p_{a|Z} = \frac{p_{Z|a} \cdot p_a}{p_Z} \tag{4.1}$$

the maximum a posteriori equation is

$$\frac{\partial \ell n p_{Z|a}}{\partial a} + \frac{\partial \ell n p_a}{\partial a} = 0 \ . \tag{4.2}$$

The classical maximum likelihood criterion is to choose that a for which $p_{Z|a}$ is a maximum. The maximum likelihood equation is then

$$\frac{\partial \ell n p_{Z|a}}{\partial a} = 0 \qquad (4.3)$$

which is the same as the maximum a posteriori criterion with no prior knowledge of the parameters.

In the next two sections the maximum a posteriori criterion is applied to our linear system with two idealized error sources: (1) white gaussian measurement noise of the output, and (2) white gaussian process noise.

B. CRITERION WITH MEASUREMENT NOISE

Without process noise and with perfect measurements of the input, u_i, the discrete system

$$x_{i+1} = \Phi x_1 + \Gamma u_i, \qquad x_o \text{ given} \qquad (4.4)$$

with measurements

$$z_i = Hx_i + v_i$$

where

$$Ev_i \, v_j^T = R_1 \, \delta_{ij} \, . \qquad (4.5)$$

The probability density of each measurement given the unknown parameters (including x_o) and the sequence u_i is gaussian:

$$p_{z_i|a} = \frac{1}{\sqrt{(2\pi)^m |R_1|}} \exp\left\{- \frac{1}{2}(z_i - Hx_i)^T R_1^{-1} (z_i - Hx_i)\right\} . \qquad (4.6)$$

Since the sequence x_i may be calculated deterministically, each measurement is independent and we may write

$$p_{Z|a} = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^m |R_1|}} \exp\left\{- \frac{1}{2}(z_i - Hx_i)^T R_1^{-1} (z_i - Hx_i)\right\} \qquad (4.7)$$

or

$$\ell n p_{Z|a} = \sum_{i=1}^{N} - \frac{1}{2} \ell n(2\pi)^m |R_1| - \tag{4.8}$$
$$- \frac{1}{2}(z_i - Hx_i)^T R_1^{-1} (z_i - Hx_i) .$$

Thus, maximizing $p_{Z|a}$ with respect to a is equivalent to minimizing the performance index

$$J = \frac{1}{2} \sum_{i=1}^{N} \left[\ell n|R_1| + (z_i - Hx_i)^T R_1^{-1} (z_i - Hx_i) \right] \tag{4.9}$$

with respect to a, subject to the constraint

$$x_{i+1} = \phi x_i + \Gamma u_i . \tag{4.10}$$

If none of the parameters in R_1 are known, then we can first minimize with respect to the parameters in R_1 to obtain [SP-1, p. 23]

$$\hat{R}_1 = \frac{1}{N} \sum_{i=1}^{N} (z_i - Hx_i)(z_i - Hx_i)^T \tag{4.11}$$

so that minimizing the performance index, (4.9) is then equivalent to minimizing

$$J = \det \left[\sum_{i=1}^{N} (z_i - Hx_i)(z_i - Hx_i)^T \right] \tag{4.12}$$

with respect to all unknown parameters except those in R_1. However, if all the parameters in R_1 are already known, then minimizing (4.9) is equivalent to minimizing

$$J = \frac{1}{2} \sum_{i=1}^{N} (z_i - Hx_i)^T R_1^{-1} (z_i - Hx_i) . \tag{4.13}$$

In the continuous system, Eq. 1.1, the assumption that the measurement noise v is white (uncorrelated) is a useful approximation if the correlation times of the measurement noise are short with respect to the dynamics of the system being measured. However, in trying to estimate the intensity matrix R, the assumption about independent measurement errors is invalid as the measurement interval tends to zero. This is reflected in the fact that the limit of (4.9) does not exist. However, we can estimate R by thinking of v as a correlated process with a very short (but finite) correlation time. In this case an estimate of R is given by

$$\hat{R} \cong \int_{-T}^{+T} C(\tau)\, d\tau \tag{4.14}$$

where the correlation matrix $C(\tau)$ is given by

$$C(\tau) = \frac{1}{T-\tau} \int_{0}^{T-\tau} v(t) v^T(t+\tau)\, dt \,. \tag{4.15}$$

The value of R is a measure of the noise characteristics of the instrumentation, and may be obtained from measuring the instrumentation alone, without exciting the system. For the remainder of this thesis, R will be assumed known. With R known, we can minimize the limit of (4.13) with $R_1 = R/\Delta t$:

$$J = \frac{1}{2} \int_{t_o}^{t_f} (z - Hx)^T R^{-1} (z - Hx)\, dt \,. \tag{4.16}$$

We are now subject to the constraint

$$\dot{x} = Fx + Gu, \qquad x(t_o) = x_o \,. \tag{4.17}$$

The latter performance index can also be derived by maximizing the likelihood ratio [ME-2]

$$L = \frac{p_{Z|H_1,a}}{p_{Z|H_o}} \,, \tag{4.18}$$

where H_1 represents the hypothesis that

$$z = Hx + v$$

and H_o represents the hypothesis that

$$z = v .$$

The criterion developed in this section is also known as the output error criterion [DE-2, ME-2]

C. CRITERION WITH MEASUREMENT AND PROCESS NOISE

With process noise, the discrete system (4.4) becomes

$$x_{i+1} = \Phi x_i + \Gamma u_i + w_i, \qquad x_o \text{ given} . \tag{4.19}$$

In calculating the correlation $E(z_i - \bar{z}_i)(z_j - \bar{z}_j)^T$ for $i \neq j$, we reduce the calculation to finding

$$M_i \triangleq E(x_i - \bar{x}_i)(x_i - \bar{x}_i)^T .$$

Refer, for the moment, to the first equation in (4.22) where $M_o = 0$ since x_o is given. For the case without process noise $Q_i = 0$, so that $M_i = 0$ and the measurements are uncorrelated. However, with process noise $Q_i \neq 0$, so that $M_i \neq 0$, and the measurements are correlated. Since the measurements are not independent, the probability density $p_{Z|a}$ cannot be equated to the product of the individual probability densities. For this reason, a Kalman filter representation is used [ME-2]. Since it is known that the "innovations" are white and contain all the statistical information contained in the measurements [KA-1], the probability density $p_{Z|a}$ is given by

$$p_{Z|a} = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^m |B_i|}} \exp\left\{-\frac{1}{2}\nu_i^T B_i^{-1} \nu_i\right\} \qquad (4.20)$$

where

$$\left.\begin{aligned} \bar{x}_{i+1} &= \phi_i \hat{x}_i + \Gamma_i u_i \\ \hat{x}_i &= \bar{x}_i + P_i H_i^T R_i^{-1}(z_i - H\bar{x}_i) \end{aligned}\right\} x_o \text{ given;} \qquad (4.21)$$

$$\left.\begin{aligned} M_{i+1} &= \phi_i P_i \phi_i^T + Q_i \\ P_i &= M_i - M_i H^T (H_i M_i H_i^T + R_i)^{-1} H_i M_i \end{aligned}\right\} M_o = 0; \qquad (4.22)$$

and

$$\nu_i = z_i - H\bar{x}_i$$

called the innovations sequence is purely random with correlation

$$\begin{aligned} B_i \delta_{ij} &= E\nu_i \nu_j^T = E[H(x_i - \bar{x}_i) + v_i][H(x_j - \bar{x}_j) + v_j]^T \\ &= (HM_i H^T + R_i)\,\delta_{ij} \,. \end{aligned} \qquad (4.24)$$

Taking the natural logarithm of (4.19), we obtain

$$\ell n p_{Z|a} = \sum_{i=1}^{N} -\frac{1}{2}\ell n(2\pi)^m |B_i| - \frac{1}{2}(z_i - H\bar{x}_i)^T B_i^{-1}(z_i - H\bar{x}_i) \,. \qquad (4.25)$$

The maximum likelihood estimate is then given by minimizing the objective function

$$J = \frac{1}{2}\sum_{i=1}^{N} \ell n|B_i| + (z_i - H\bar{x}_i)^T B_i^{-1}(z_i - H\bar{x}_i) \qquad (4.26)$$

with respect to the vector a of unknown parameters in ϕ, Γ, H, Q_1, R_1, and x_o subject to the two constraint equations

$$\bar{x}_{i+1} = \phi\bar{x}_i + \Gamma u + \phi\left[M_i - M_iH^T(HM_iH^T+R_1)^{-1}HM_i\right](H^TR^{-1}(z_i-Hx_i)$$

$$\bar{x}_o = x_o \ ; \tag{4.27}$$

$$M_{i+1} = \phi\left[M_i - M_iH^T(HM_iH^T + R_1)^{-1}HM_i\right]\phi + Q_1, \quad M_o = 0 \ .$$

If we can make the assumption that M_i is a constant, then considerable simplification results. This will eliminate the second set of constraint equations in (4.27). This assumption will be a good one if the test is conducted over a long time interval so that M_i is nearly constant for most of the test. However, if this assumption is not valid, then we must solve the problem as formulated above.

In the "steady state Kalman filter representation" [ME-2], we can identify B and K instead of R_1 and Q_1 where B and K are given by $B = HMH^T + R_1$, and $K = MH^TB^{-1}$ and M is the solution to

$$M = \phi[M - MH^T(HMH^T + R_1)^{-1}HM]\phi + Q_1 \ .$$

Note that the above equations cannot be solved uniquely for Q_1. Our problem now becomes: minimize the performance index

$$J = \frac{1}{2}\sum_{i=1}^{N}\left[\ln|B| + (z_i - Hx_i)^TB^{-1}(z_i - Hx_i)\right] \tag{4.28}$$

with respect to the parameters in ϕ, Γ, H, B, K, and x_o, subject to the constraint

$$x_{i+1} = \phi x_i + \Gamma u_i + \phi K(z_i - Hx_i) \ . \tag{4.29}$$

For the continuous case we can proceed in a similar manner. If we assume that R is known, then the identification criterion for the steady state Kalman-Bucy filter representation is to minimize

$$J = \frac{1}{2}\int_{t_o}^{t_f} (z - H\hat{x})^T R^{-1} (z - H\hat{x}) dt \tag{4.30}$$

with respect to the unknown parameters in F, G, H, K, and x_o subject to the constraint

$$\dot{\hat{x}} = F\hat{x} + Gu + K(z - H\hat{x}), \qquad \hat{x}(t_o) = x_o \; . \tag{4.31}$$

As in the discrete case, if the assumptions regarding the steady state are not valid, then we must include the covariance equation as another constraint.

This criterion could also be derived by employing the criterion for the maximum likelihood estimate of a and the trajectory $x(t)$, $t_o \leq t \leq t_f$. In this case we want to minimize

$$J = \frac{1}{2}[x(t_o)-x_o]^T P_o^{-1} [x_o(t_o) - x_o] + \frac{1}{2}\int_{t_o}^{t_f} \left[w^T Q^{-1} w + (z - Hx)^T R^{-1} (z - Hx) \right] dt \tag{4.32}$$

with respect to a and $w(t)$, $t_o \leq t \leq t_f$; subject to

$$\dot{x} = Fx + Gu + w \; . \tag{4.33}$$

By performing the minimization first with respect to $w(t)$, we obtain the Kalman-Bucy filter equations

$$\begin{aligned} \dot{\hat{x}} &= F\hat{x} + Gu + PH^T R^{-1} (z - H\hat{x}), & \hat{x}(t_o) &= x_o \\ \dot{P} &= FP + PF^T + Q - PH^T R^{-1} HP, & P(t_o) &= P_o \end{aligned} \tag{4.34}$$

and the equation for the adjoint variable

$$\dot{\lambda} = -(F - PH^TR^{-1}H)^T\lambda + H^TR^{-1}(z - Hx), \qquad \lambda(t_f) = 0 . \qquad (4.35)$$

If we substitute $w = -QG^T\lambda$ and $x = \hat{x} - P\lambda$ into (4.32), and add the differential

$$-\frac{d}{dt}\left\{\lambda^T P\lambda\right\}$$

inside the integral and

$$\lambda^T(t_f)\ P(t_f)\ \lambda(t_f) - \lambda^T(t_o)\ P(t_o)\ \lambda(t_o)$$

outside the integral, we obtain (4.30). Our identification criterion then is to minimize (4.30) with respect to a, subject to (4.34). The adjoint equation (4.35) is not considered a constraint for the minimization with respect to a since λ is not in (4.30) or (4.34). Once the maximum a posteriori estimate of a has been found, the smoothed estimate of the trajectory using $a = \hat{a}$ is the maximum a posteriori estimate of the trajectory.

If we assume perfect measurements of the state and derivatives of the state are taken, then the criterion of (4.32) and (4.33) may be reduced to minimizing

$$J = \frac{1}{2}\int_{t_o}^{t_f} (\dot{x} - Fx - Gu)^T Q^{-1}(\dot{x} - Fx - Gu)dt$$

with respect to the unknowns in F and G. Since the unknown parameters in F and G are quadratic in (4.36), estimates may be obtained in one step. This criterion is a special case of the criterion discussed in this section and is known as the equation-error criterion [DE-2 and ME-2].

D. CRITERION WITH PRIOR INFORMATION

To incorporate prior information, let us use the maximum a posteriori equation and assume a prior probability distribution that is gaussian with mean $\bar{a}$ and covariance A:

$$p_a = \frac{1}{\sqrt{(2\pi)^q |A|}} \exp\left\{-\frac{1}{2}(a-\bar{a})^T A^{-1} (a-\bar{a})\right\} \quad (4.37)$$

or

$$\ln p_a = -\frac{1}{2} \ln (2\pi)^q |A| - \frac{1}{2}(a-\bar{a})^T A^{-1} (a-\bar{a}) \; . \quad (4.38)$$

The performance indices are then modified to include the additional term

$$\frac{1}{2} (a - \bar{a})^T A^{-1} (a - \bar{a})$$

and the constraint equations remain the same.

Chapter V

IDENTIFICATION ALGORITHMS

A. QUASILINEARIZATION*

Denery [DE-1] combines two different linearization techniques to minimize the output error performance index

$$J = \frac{1}{2}\int_{t_o}^{t_f} (z - \hat{z})^T R^{-1}(z - \hat{z})dt \tag{5.1}$$

where the system is modelled by

$$\begin{aligned} \dot{\hat{x}} &= F\hat{x} + Gu, \qquad \hat{x}(t_o) = x_o \\ \hat{z} &= H\hat{x} \end{aligned} \tag{5.2}$$

and z is a given set of measurements. J is minimized with respect to the unknown parameters in F, G, H, and x_o, subject to the constraints in (5.2). His first linearization technique may be considered an extension of quasilinearization. Instead of modelling the system as given by (5.2), $\hat{z}$ is instead modelled by

$$\begin{aligned} \dot{\hat{x}} &= F\hat{x} + Gu + D(z-H\hat{x}) = F_n\hat{x} + Gu + Dz, \qquad \hat{x}(t_o) = x_o \\ \hat{z} &= H\hat{x} + L(z - H\hat{x})' = H_N\hat{x} + Lz \end{aligned} \tag{5.3}$$

where

* Denery's combined algorithm [DE-1].

$$F_N \triangleq F - DH$$
$$H_N \triangleq H - LH \ . \tag{5.4}$$

This set of equations is useful only if the system (2) is in a Denery canonical form. Now, let

$$\delta G = G - G_N$$
$$\delta x_o = x_o - x_{N_o} \tag{5.5}$$

and define z_N by

$$\dot{x}_N = F_N x_N + G_N u, \qquad x_N(t_o) = x_{N_o}$$
$$z_N = H_N x_N \ . \tag{5.6}$$

If we guess F_N, G_N, H_N, and x_{N_o}, the unknown parameters are now in D, δG, L, and δx_o instead of F, G, H, and x_o. By augmenting the system equations with the terms $D(z-H\hat{x})$ and $L(z-H\hat{x})$, Denery was able to make the unknown parameters coefficients of known functions so that we may write

$$\hat{z} = z_N + \left(\frac{\partial z}{\partial \alpha}\right)\alpha \tag{5.7}$$

where α is a $(q \times 1)$ vector representing the unknown parameters in δG, δx_o, D, and L. The ith column of the matrix $(\partial z/\partial \alpha)$ is given by the sensitivity equations

$$\dot{\left(\frac{\partial \hat{x}}{\partial \alpha_i}\right)} = F_N\left(\frac{\partial \hat{x}}{\partial \alpha_i}\right) + \frac{\partial D}{\partial \alpha_i} z + \left(\frac{\partial \delta G}{\partial \alpha_i}\right)u, \qquad \frac{\partial \hat{x}}{\partial \alpha_i}(t_o) = \frac{\partial \delta x_o}{\partial \alpha_i}$$
$$\left(\frac{\partial \hat{z}}{\partial \alpha_i}\right) = H_N\left(\frac{\partial \hat{x}}{\partial \alpha_i}\right) + \frac{\partial L}{\partial \alpha_i} z \ . \tag{5.8}$$

Notice that in this formulation, the sensitivity equations are driven by the actual measurements z. Taking the derivative of the performance index with respect to the unknown parameters,

$$\frac{\partial J}{\partial \alpha} = \int_{t_o}^{t_f} (z - \hat{z})^T R^{-1} (-) \frac{\partial \hat{z}}{\partial \alpha} dt = 0$$

and substituting (5.7) into the result yields

$$\int_{t_o}^{t_f} \left(\frac{\partial \hat{z}}{\partial \alpha}\right)^T R^{-1} \left(\frac{\partial \hat{z}}{\partial \alpha}\right) \alpha \, dt = \int_{t_o}^{t_f} \left(\frac{\partial \hat{z}}{\partial \alpha}\right)^T R^{-1} (z - z_N) dt$$

so that an estimate of α is given by

$$\hat{\alpha} = \left[\int_{t_o}^{t_f} \left(\frac{\partial \hat{z}}{\partial \alpha}\right)^T R^{-1} \left(\frac{\partial \hat{z}}{\partial \alpha}\right) dt\right]^{-1} \left[\int_{t_o}^{t_f} \left(\frac{\partial \hat{z}}{\partial \alpha}\right)^T R^{-1} (z - z_N) dt\right] . \tag{5.10}$$

An estimate of the unknowns in F, G, H, and x_o is then given by employing (5.4) and (5.5):

$$\begin{aligned} \hat{G} &= G_N + \hat{\delta G} \\ \hat{x}_o &= x_{N_o} + \hat{\delta x}_o \\ \hat{H} &= (I - \hat{L})^{-1} H_N \\ \hat{F} &= F_N + \hat{D}\hat{H} . \end{aligned} \tag{5.11}$$

These estimates may then be used as nominal values in another iteration.

This approach was found to be convergent even for large inaccuracies in the initial guesses of the unknown parameters. However, the estimates given by this method are biased even if the noise is unbiased (i.e., has zero mean value).

After three or four iterations of using this extended quasilinearization technique, Denery suggests switching to the normal quasilinearization technique. In the method of quasilinearization, z is represented by (5.2) but approximated by small deviations from the nominal by

$$\hat{z} = H_N x_N + (H - H_N)x + H_N \delta x = H_N x_A + (H-H_N)x \tag{5.12}$$

where $x_A \triangleq x_N + \delta x$, x_N is given by (5.6), and δx is determined from

$$\delta\dot{x} = F_N \delta x + (F-F_N)x + \delta G u, \qquad \delta x(t_o) = \delta x_o \tag{5.13}$$

so that x_A is determined from

$$\dot{x}_A = F_N x_A + (F - F_N)x_N + (G_N + \delta G)u, \qquad x_A(t_o) = x_{N_o} + \delta x_o \ . \tag{5.14}$$

For quasilinearization, we assume that $F - F_N$ and $H - H_N$ are small so that for a system in a Denery canonical form, we may write

$$\begin{aligned} F - F_N &= DH = D(H_N + LH) \approx DH_N \\ H - H_N &= LH = L(H_N + LH) \approx LH_N \end{aligned} \tag{5.15}$$

where D and L are small. Now substituting these into (5.14), we have

$$\begin{aligned} \dot{x}_A &= F_N x_A + Dz_N + (G_N+\delta G)u, \quad x_A(t_o) = x_{N_o} + \delta x_o \\ \hat{z} &= H_N x_A + Lz_N \ . \end{aligned} \tag{5.16}$$

This equation is identical to (5.3) except that z_N has replaced z. The solution is the same as the extended method except that z_N drives the sensitivity equations (5.8) instead of z.

The estimates obtained using this method are unbiased but the method often does not converge if the initial guesses of the unknown parameters are far from their true values. Thus, it can be used after the first method to obtain a combined algorithm insensitive to inaccuracies in the initial values of the unknown parameters and yielding an unbiased estimate.

In summary, to find an estimate for F, G, H, and x_o with initial guesses given by F_N, G_N, H_N, and x_{N_o}:

1. Calculate a nominal trajectory

$$\dot{x}_N = F_N x_N + G_N u, \qquad x_N(t_o) = x_{N_o}$$
$$z_N = H_N x_N \;. \qquad \text{per (5.6)}$$

2. Calculate the sensitivity functions given by z or z_N

$$\dot{\left(\frac{\partial \hat{x}}{\partial \alpha_i}\right)} = F_N\left(\frac{\partial \hat{x}}{\partial \alpha_i}\right) + \frac{\partial D}{\partial \alpha_i} z_{(n)} + \left(\frac{\partial \delta G}{\partial \alpha_i}\right)u, \qquad \left(\frac{\partial \hat{x}}{\partial \alpha_i}\right)(t_o) = \frac{\partial \delta x_o}{\partial \alpha_i}$$
$$\frac{\partial \hat{z}}{\partial \alpha_i} = H_N\left(\frac{\partial \hat{x}}{\partial \alpha_i}\right) + \frac{\partial L}{\partial \alpha_i} z_{(n)}\,, \qquad i = 1, 2, \ldots q \;. \qquad \text{per (5.8)}$$

3. Calculate an estimate of the unknown parameters in δG, $\delta\hat{x}_o$, D, and L:

$$\hat{\alpha} = \left[\int_{t_o}^{t_f} \left(\frac{\partial \hat{z}}{\partial \alpha}\right)^T R^{-1} \left(\frac{\partial \hat{z}}{\partial \alpha}\right) dt\right]^{-1} \left[\int_{t_o}^{t_f} \left(\frac{\partial \hat{z}}{\partial \alpha}\right)^T R^{-1} (z - z_N)\, dt\right]. \qquad \text{per (5.10)}$$

4. Calculate estimates of the parameters that can be used as nominal values in the next iteration

$$\hat{G} = G_N + \hat{\delta G}$$
$$\hat{x}_o = x_{N_o} + \hat{\delta x}_o$$
$$\hat{H} = (I - \hat{L})^{-1} H_N$$
$$\hat{F} = F_N + \hat{D}\hat{H} \;. \qquad \text{per (5.11)}$$

The amount of computation per iteration involves $n + n \cdot q + \frac{1}{2}q(q+1) + q$ integrations over the length of the test.

Example: Identify the constant, a, in the first order system:

$$\dot{x} = -ax + au, \qquad x(0) = 0$$

$$z = x + v$$

where $Ev(t)v(\tau) = r\delta(t - \tau)$. Note that this example is slightly different from our development, since the same parameter is in F and G. Augmenting the state equation with $D(z-\hat{x})$, we have

$$\dot{\hat{x}} = -a\hat{x} + au + D(z-\hat{x}) = -(a+D)\hat{x} + au + Dz \; .$$

Now,

$$\delta G = G - G_N = a - a_N = -D \; .$$

Let

$$\delta G = \alpha \text{ so that } \frac{\partial \delta G}{\partial \alpha} = 1 \text{ and } \frac{\partial D}{\partial \alpha} = -1 \; .$$

The nominal and sensitivity equations are

$$\dot{x}_N = -a_N x_N + a_N u, \qquad x_N(o) = 0$$

$$\dot{\left(\frac{\partial \hat{x}}{\partial \alpha}\right)} = -a_N\left(\frac{\partial \hat{x}}{\partial \alpha}\right) + u - z, \qquad \frac{\partial \hat{x}}{\partial \alpha}(o) = 0 \; .$$

An estimate of α is given by

$$\hat{\alpha} = \left[\int_o^T \left(\frac{\partial \hat{x}}{\partial \alpha}\right)^2 dt\right]^{-1} \int_o^T \left(\frac{\partial \hat{x}}{\partial \alpha}\right)(z - x_N)dt \; .$$

An updated estimate of a (which can be used as a nominal value for the next iteration) is given by

$$\hat{a} = a_N + \hat{\alpha} .$$

For the second set of iterations, the only change is that x_N replaces z in the sensitivity equation.

B. PROCESS NOISE

With process noise, we can represent system Eq. (5.2) by its steady state Kalman filter

$$\begin{aligned} \dot{\hat{x}} &= F\hat{x} + Gu + K(z - H\hat{x}) \\ \hat{z} &= H\hat{x} . \end{aligned} \tag{5.17}$$

If we proceed as before with Denery's extension, we replace (5.17) with

$$\begin{aligned} \dot{\hat{x}} &= F\hat{x} + Gu + K(z - H\hat{x}) + D(z - H\hat{x}) \\ \hat{z} &= H\hat{x} + L(z - H\hat{x}) . \end{aligned} \tag{5.18}$$

Obviously the sum $K + D$ may be identified by Denery's extension but K and D cannot be identified separately. However, we can identify F, G, H, and x_o by the first quasilinearization technique, assuming that $K \approx 0$ and proceed to the second technique.

Proceeding with the second quasilinearization method we approximate $\hat{z}$ in (5.17) with

$$\hat{z} = H_N x_N + (H-H_N)x_N + H_N \delta x = H_N x_A + (H-H_N)x_N \tag{5.19}$$

where $x_A = x_N + \delta x$, and x_N and δx are given by

$$\begin{aligned} \dot{x}_N &= F_N x_N + G_N u + K_N(z-H_N x_N), \qquad x_N(t_o) = x_{N_o} \\ z_N &= H_N x_N \end{aligned} \tag{5.20}$$

and

C

$$\delta\dot{x} = F_N \delta x + (F-F_N)x + (G-G_N)u + (K-K_N)(z-H_N x_N)$$
$$+ K_N[-(H-H_N)x_N - H_N \delta x], \qquad \delta x(t_o) = \delta x_o \tag{5.21}$$

so that x_A is given by

$$\dot{x}_A = F_N x_A + \delta F x_N + (G_N+\delta G)u + (K_N+\delta K)(z-H_N x_N)$$
$$+ K_N[(H_N-\delta H)x_N - H_N x_A], \qquad x_A(t_o) = x_{N_o} + \delta x_o . \tag{5.22}$$

For a Denery canonical form we can write

$$\delta F = F - F_N = DH = D(H_N+LH) \approx DH_N$$
$$\delta H = H - H_N = LH = L(H_N+LH) \approx LH_N . \tag{5.23}$$

Substituting and simplifying, we have

$$\dot{x}_A = (F_N-K_N H_N)x_A + (G_N+\delta G)u + Dz_N + K_N z + \delta K(z-z_N)$$
$$- K_N L z_N, \qquad x_A(t_o) = x_{N_o} + \delta x_o \tag{5.24}$$

$$\hat{z} = H_N x_A + L z_N .$$

Let α represent the unknown parameters in δG, δx_o, D, L, and δK. The sensitivity equations become

$$\left(\frac{\partial \dot{x}_A}{\partial \alpha_i}\right) = (F_N-K_N H_N)\left(\frac{\partial x_A}{\partial \alpha_i}\right) + \frac{\partial D}{\partial \alpha_i} z_N + \frac{\partial \delta G}{\partial \alpha_i} u + \frac{\partial \delta K}{\partial \alpha_i}(z-z_N) - \tag{5.25}$$
$$- K_N \frac{\partial L}{\partial \alpha_i}(z - z_N) ,$$

$$\frac{\partial x_A}{\partial \alpha_i}(t_o) = \frac{\partial \delta x_o}{\partial \alpha_i}$$

$$\left(\frac{\partial \hat{z}}{\partial \alpha_i}\right) = H_N\left(\frac{\partial x_A}{\partial \alpha_i}\right) + \frac{\partial L}{\partial \alpha_i} z_N \; . \qquad (5.25) \text{ Cont.}$$

Note that these sensitivity equations are driven by both z and z_N. An estimate of α is given by (5.10) where $(\partial\hat{z}/\partial\alpha)$ is now given by (5.25) and estimates of F, G, H, and x_o are given by (5.11). An estimate of K is given by

$$\hat{K} = K_N + \hat{\delta K} \; . \qquad (5.26)$$

Example. Identify a and K for the first order system

$$\dot{x} = -ax + au + w , \qquad x(0) = 0$$

$$z = x + v$$

and its steady state Kalman Filter representation

$$\dot{\hat{x}} = -a\hat{x} + au + K(z - \hat{x}), \qquad \hat{x}(0) = 0 \; .$$

For the first part of the algorithm, use the same algorithm as the previous example, assuming that $K = 0$. For the second part, the nominal trajectory is given by

$$\dot{x}_N = -a_N x_N + a_N u + K_N(z - x_N), \qquad x_N(0) = 0$$

where for the first iteration, $K_N = 0$, and a_N equals its identified value from the first part of the algorithm. The approximate trajectory is given by

$$\dot{x}_A = -(a_N + K_N)x_A + (a_N + \delta K)u + Dx_N + K_N z + \delta K(z - x_N)$$

$$x_A(0) = 0$$

$$\hat{z} \approx x_A .$$

Let $\alpha_1 = \delta G = -D$ and $\alpha_2 = \delta K$. The sensitivity equations for α_1 and α_2 are

$$\left(\frac{\partial \dot{x}_A}{\partial \alpha_1}\right) = -(a_N + K_N)\left(\frac{\partial x_A}{\partial \alpha_1}\right) + u - x_N, \qquad \frac{\partial x_A}{\partial \alpha_1}(0) = 0$$

$$\left(\frac{\partial \dot{x}_A}{\partial \alpha_2}\right) = -(a_N + K_N)\left(\frac{\partial x_A}{\partial \alpha_2}\right) + z - x_N, \qquad \frac{\partial x_A}{\partial \alpha_2}(0) = 0 .$$

Estimates of α_1 and α_2 are given by

$$\begin{pmatrix} \hat{\alpha}_1 \\ \\ \hat{\alpha}_2 \end{pmatrix} = \begin{bmatrix} \int_0^T \left(\frac{\partial x_A}{\partial \alpha_1}\right)^2 dt & \int_0^T \left(\frac{\partial x_A}{\partial \alpha_1}\right)\left(\frac{\partial x_A}{\partial \alpha_2}\right) dt \\ \int_0^T \left(\frac{\partial x_A}{\partial \alpha_1}\right)\left(\frac{\partial x_A}{\partial \alpha_2}\right) dt & \int_0^T \left(\frac{\partial x_A}{\partial \alpha_2}\right)^2 dt \end{bmatrix}^{-1} \begin{bmatrix} \int_0^T \left(\frac{\partial x_A}{\partial \alpha_1}\right)(z - x_N)\, dt \\ \int_0^T \left(\frac{\partial x_A}{\partial \alpha_2}\right)(z - x_N)\, dt \end{bmatrix} .$$

Updated estimates for a and K are given by

$$\hat{a} = a_N + \hat{\alpha}_1$$

$$\hat{K} = K_N + \hat{\alpha}_2 .$$

The only problem in implementing this algorithm is that the term

$$\int_0^T \left(\frac{\partial x_A}{\partial \alpha_2}\right)^2 dt$$

may be too small to allow an accurate estimate of K and the algorithm will not converge. For $x_N \approx x$, the second sensitivity equation takes the form

$$\dot{x} = -ax + v, \qquad x(0) = 0 \; ,$$

so that $P = E(x^2)$ is given by

$$\dot{P} = -2aP + r, \qquad P(0) = 0 \; ,$$

or

$$P = \frac{r}{2a}\left(1 - e^{-2at}\right) .$$

Actually, to be consistent with our steady-state Kalman filter hypothesis of a long test, we may set

$$E(x^2) = \frac{r}{2a} .$$

The covariance of K (assuming a is known perfectly) is given by

$$P_K = \left[\frac{1}{r}\int_o^T \frac{r}{2a}\, dt\right]^{-1} = \frac{2(a_N + K_N)}{T} ,$$

so that no matter what the input is we must have a sufficiently long test to estimate K.

C. GRADIENT METHODS*

Minimize the output error performance index

$$J = \frac{1}{2}\int_{t_o}^{t_f} (z - Hx)^T R^{-1}(z - Hx)\,dt \tag{5.26}$$

* First paragraph based on Sage and Melsa [SA-2].

subject to the constraints

$$\dot{x} = Fx + Gu, \qquad x(t_o) = x_o$$
$$\dot{a}' = 0 \tag{5.27}$$

where a' is a $q' \times 1$ vector that denotes those unknown parameters in F, G, and H. The Hamiltonian is

$$\mathcal{H} = \frac{1}{2}(z - Hx)^T R^{-1}(z - Hx) + \lambda^T(Fx + Gu) + \Gamma \cdot 0 , \tag{5.28}$$

where λ and Γ are conjugate to x and a'. The adjoint equations are given by

$$\dot{\lambda}^T = \frac{\partial \mathcal{H}}{\partial x} = (z-Hx)^T R^{-1} H - \lambda^T F, \qquad \lambda^T(t_f) = 0$$

$$\dot{\Gamma}_i = -\frac{\partial \mathcal{H}}{\partial a_i} = (z - Hx)^T R^{-1} \frac{\partial H}{\partial a_i} x - \lambda^T\left(\frac{\partial F}{\partial a_i} x + \frac{\partial G}{\partial a_i} u\right) \tag{5.29}$$

$$\Gamma_i(t_f) = 0,$$

$$i = 1, 2, \ldots q' .$$

u and z are given functions so that the Hamiltonian is not minimized with respect to u. The gradients with respect to a' and x are given by

$$\Gamma(t_o) = \frac{\partial J}{\partial a'} \quad \text{and} \quad \lambda(0) = \frac{\partial J}{\partial x(t_o)} . \tag{5.30}$$

A steepest descent or conjugate gradient algorithm can now be implemented as follows:

(1) Guess an initial value for a' and x_o;

(2) Calculate x by integrating (per Eq. 5.27),

$$\dot{x} = Fx + Gu, \qquad x(t_o) = x_o .$$

(3) Calculate the adjoint equations (per Eq. 5.29)

$$\dot{\lambda} = H^T R^{-1}(z - Hx) - F^T\lambda, \qquad \lambda(t_f) = 0$$

$$\dot{\Gamma}_i = x^T \frac{\partial H^T}{\partial a_i} R^{-1}(z-Hx) - \left[x^T \frac{\partial F^T}{\partial a_i} + u^T \frac{\partial G^T}{\partial a_i}\right]\lambda, \qquad \Gamma_i(t_f) = 0 .$$

(4) Values of a' and $x(0)$ are updated according to

$$\begin{aligned} a'^{new} &= a'^{old} - K\Gamma(0) \\ x_o^{new} &= x_o^{old} - K\lambda(0) \end{aligned} \tag{5.31}$$

for the steepest descent algorithm and in a conjugate direction for the conjugate gradient algorithm. This approach requires integrating $n + n + q'$ first order differential equations over the length of the test.

Another approach is to take the derivative of J directly:

$$\left.\frac{\partial J}{\partial a}\right|_{a=a^{old}} = \int_{t_o}^{t_f} (z-Hx)^T R^{-1}(-) \left[\frac{\partial H}{\partial a} x + H \frac{\partial x}{\partial a}\right]dt \tag{5.32}$$

where a (not a') represents unknowns in x_o as well as F, G, and H. $(\partial x/\partial a_i)$ is generated by the sensitivity equation

$$\begin{aligned} \dot{\left(\frac{\partial x}{\partial a_i}\right)} &= F\left(\frac{\partial x}{\partial a_i}\right) + \frac{\partial F}{\partial a_i} x + \frac{\partial G}{\partial a_i} u \\ \frac{\partial x}{\partial a_i}(t_o) &= \frac{\partial x_o}{\partial a_i} \\ i &= 1, 2, \ldots q . \end{aligned} \tag{5.33}$$

This version of the algorithm may be implemented as above except steps (3) and (4) are replaced by (3') and (4').

(3'). For each a_i, calculate $(\partial x/\partial a_i)$ and $(\partial J/\partial a_i)$ according to

$$\frac{\partial J}{\partial a_i} = \int_{t_o}^{t_f} (z - Hx)^T R^{-1} (-) \left[\frac{\partial H}{\partial a_i} x + H\left(\frac{\partial x}{\partial a_i}\right)\right] dt \tag{5.34}$$

$$\left(\frac{\partial \dot{x}}{\partial a_i}\right) = F\left(\frac{\partial x}{\partial a_i}\right) + \frac{\partial F}{\partial a_i} x + \frac{\partial G}{\partial a_i} u, \quad \frac{\partial x}{\partial a_i}(t_o) = \frac{\partial x_o}{\partial a_i}, \tag{5.35}$$

$$i = 1, 2, \ldots q .$$

(4'). Values of a are updated according to

$$a_i^{new} = a_i^{old} - K\left(\frac{\partial J}{\partial a_i}\right) \tag{5.36}$$

for the steepest descent algorithm and in a conjugate direction for the conjugate gradient algorithm. This approach requires more computation and will not be considered further.

With the addition of process noise, our original algorithm remains valid except that (5.27) is replaced by

$$\dot{\hat{x}} = F\hat{x} + Gu + K(z - H\hat{x}), \qquad \hat{x}(t_o) = x_o \tag{5.37}$$

and the adjoint equations, (5.29), are replaced by

$$\dot{\lambda} = H^T R^{-1}(z - H\hat{x}) - (F - KH)^T\lambda, \qquad \lambda(t_f) = 0$$

$$\dot{\Gamma}_i = \hat{x}^T \frac{\partial H^T}{\partial a_i} R^{-1}(z-H\hat{x}) - \tag{5.38}$$

$$- \left[\hat{x}^T \frac{\partial F^T}{\partial a_i} + u^T \frac{\partial G^T}{\partial a_i} + (z-H\hat{x})^T \frac{\partial K^T}{\partial a_i} - \hat{x}^T \frac{\partial H^T}{\partial a_i} K^T\right]\lambda,$$

$$\Gamma_i(t_f) = 0, \qquad i = 1, 2, \ldots q' .$$

Chapter VI

OPTIMAL INPUT CRITERIA

A. INTRODUCTION

If we expand one of the identification performance indices of Chapter IV to second order in a, we have

$$J(a) = J(\hat{a}) + \left.\frac{\partial J}{\partial a}\right|_{a=\hat{a}} (a-\hat{a}) + \frac{1}{2}(a-\hat{a})^T \left.\frac{\partial^2 J}{\partial a^2}\right|_{a=\hat{a}} (a-\hat{a}) + \cdots \qquad (6.1)$$

The minimization algorithms of Chapter V satisfy the likelihood equation

$$\left.\frac{\partial J}{\partial a}\right|_{a=\hat{a}} = 0 . \qquad (6.2)$$

The matrix

$$\left.\frac{\partial^2 J}{\partial a^2}\right|_{a=\hat{a}}$$

is a function of the input. If it is maximized (in some sense), then an iterative identification algorithm will converge faster and to a more accurate result. This is our criterion for optimizing the input.

B. THE INFORMATION MATRIX

The Fisher information matrix (Chapter II, section D) corresponding to the probability distribution $p(a|Z)$ is defined as

$$I_a \triangleq -E\left\{\frac{\partial^2 \ell n p_{a|Z}}{\partial a^2}\right\} = -E\left\{\frac{\partial^2 \ell n p_a}{\partial a^2}\right\} - E\left\{\frac{\partial^2 \ell n p_{Z|a}}{\partial a^2}\right\} = E\,\frac{\partial^2 J}{\partial a^2} \tag{6.3}$$

which is the expectation of the matrix above. p_a denotes the prior probability distribution of a (without measurements). If the prior probability density is gaussian with covariance A, then we have

$$-E\left\{\frac{\partial^2 \ell n p_a}{\partial a^2}\right\} = A^{-1} . \tag{6.4}$$

The Cramer-Rao lower bound for P_a, the covariance of a, is the inverse of the Fisher information matrix, i.e.,

$$\mathrm{var}(\hat{a}_i - a_i) \geq [I_a^{-1}]_{ii} \tag{6.5}$$

and

$$P_a \geq I_a^{-1} \tag{6.6}$$

where the equality holds if and only if [VA-1, Part I]

$$\hat{a}_i - a_i = \sum_{j=1}^{q} K_{ij}(a) \frac{\partial \ell n p_{Z|a}}{\partial a_j} , \qquad i = 1, 2, \ldots q . \tag{6.6}$$

The inverse to Fisher's information matrix represents an objective function in u to minimize. Since it is only a lower bound to the covariance, we should immediately ask how "good" a lower bound it is. In simulations done by the author, it appears to be a "good" bound in that the actual covariance is close to it. (See the simulation done in Chapt. VIII.)

There are other lower bounds that should be better: (1) the Bhattacharyya lower bound which involves higher partial derivatives in $p_{a|Z}$, and, (2) the Barankin bound which provides the greatest lower bound [VA-1, BH-1, BA-1]. Since these bounds involve considerably more computation for a marginal increase in accuracy, they will not be considered further. Let us therefore assume that the approximation

$$P_a \cong I_a^{-1} \tag{6.8}$$

is valid.

If we formally take the second derivative of Eq. (4.29) with respect to a, then the i,j^{th} element of the information matrix is given by

$$I_{ij} = E\frac{\partial^2 J}{\partial a_i \partial a_j} = \int_{t_o}^{t_f}\left[\frac{\partial H}{\partial a_j}x + H\left(\frac{\partial x}{\partial a_j}\right)\right]^T R^{-1}\left[\frac{\partial H}{\partial a_i}x + H\left(\frac{\partial x}{\partial a_i}\right)\right]dt \tag{6.9}$$

where

$$\dot{x} = Fx + Gu, \qquad x(t_o) = x_o \tag{6.10}$$

and

$$\left(\dot{\frac{\partial x}{\partial a_i}}\right) = F\left(\frac{\partial x}{\partial a_i}\right) + \frac{\partial F}{\partial a_i}x + \frac{\partial G}{\partial a_i}u, \qquad \frac{\partial x}{\partial a_i}(t_o) = \frac{\partial x_o}{\partial a_i}. \tag{6.11}$$

The indirect method for calculating the information matrix is presented in the next section with the criterion determined from the gradient algorithm.

The desired accuracy in our estimate of each parameter would depend upon the purpose of our identification. For example, if we built an observer/controller designed according to our estimates, any deviation from the true values would result in an increase in the performance index. Our design may be insensitive to some of the parameters or combinations of parameters but very sensitive to others. We may therefore weight D appropriately in an input performance index

$$\phi = \mathrm{Tr}\, D I_a^{-1}. \tag{6.12}$$

In general, the magnitude of D will depend upon the unknown parameters we are trying to estimate.

Instead of minimizing the trace of I_a^{-1}, a number of authors maximize the trace of I directly [AO-1, NA-1, and ME-3). This is simpler to do since the performance index is then a quadratic function of the sensitivity functions. The problem with maximizing the diagonal elements of I directly is the possibility that off-diagonal elements become large (in relation to the diagonal elements) so that the determinant is nearly singular. In such cases, the diagonal elements of I_a^{-1} can be very large, even though the diagonal elements of I_a are small. The following simple example illustrates this danger.

Example: First order system with two unknown parameters. Find the optimal input to identify the two parameters a and b of the first order system

$$\dot{x} = -ax + bu, \qquad x(0) = 0$$

$$z = x + v$$

where

$$Ev(t)v(t') = r\delta(t - t')$$

and there is an amplitude constraint on the input

$$|u| \leq m .$$

The sensitivity equations are

$$\dot{\left(\frac{\partial x}{\partial a}\right)} = -a\left(\frac{\partial x}{\partial a}\right) - x$$

$$\dot{\left(\frac{\partial x}{\partial b}\right)} = -a\left(\frac{\partial x}{\partial b}\right) + u .$$

By amplitude and time scaling, the above equations become

$$\dot{x}_1 = -x_1 + u, \qquad x_1(0) = 0$$

$$\dot{x}_2 = -x_2 - x_1, \qquad x_2(0) = 0$$

$$\dot{x}_3 = -x_3 + u, \qquad x_3(0) = 0$$

where a dot now denotes differentiation with respect to τ and

$$x_1 \triangleq \frac{ax}{bm}$$

$$x_2 \triangleq \frac{a^2}{bm}\left(\frac{\partial x}{\partial a}\right)$$

$$x_3 \triangleq \frac{a}{m}\left(\frac{\partial x}{\partial b}\right)$$

$$\tau \triangleq at .$$

The information matrix for a and b for a test of T sec is

$$I = \frac{1}{r}\begin{bmatrix} \int_o^T \left(\frac{\partial x}{\partial a}\right)^2 dt & \int_o^T \left(\frac{\partial x}{\partial a}\right)\left(\frac{\partial x}{\partial b}\right) dt \\ \int_o^T \left(\frac{\partial x}{\partial a}\right)\left(\frac{\partial x}{\partial b}\right) dt & \int_o^T \left(\frac{\partial x}{\partial b}\right)^2 dt \end{bmatrix}$$

$$= \frac{1}{r}\begin{bmatrix} \frac{b^2m^2}{a^5} v_2 & \frac{bm^2}{a^4} v_3 \\ \frac{bm^2}{a^4} v_3 & \frac{m^2}{a^3} v_1 \end{bmatrix}$$

where v_1, v_2, and v_3 represent the quadratures

$$v_1 \triangleq \int_o^{T'} x_3^2 \, d\tau$$

$$v_2 \triangleq \int_o^{T'} x_2 x_3 \, d\tau$$

$$v_3 = \int_0^{T'} x_2^2 \, d\tau$$

and T' is the length of the test in nondimensional times units. The covariance matrix of a and b is approximated by

$$P \cong I^{-1} = r \frac{\begin{bmatrix} \frac{a^5}{b^2 m^2} v_1 & -\frac{a^4}{bm^2} v_3 \\ -\frac{a^4}{bm^2} v_3 & \frac{a^3}{m^2} v_2 \end{bmatrix}}{v_1 v_2 - v_3^2} .$$

Let us choose our input criterion as the weighted trace

$$\phi = \text{Tr}\, DP = \text{Tr}\, \bar{D}\bar{P}$$

where $\bar{D}$ and $\bar{P}$ represent the normalized weighting and covariance matrices

$$\bar{D} = r \begin{bmatrix} \frac{a^5}{b^2 m^2} D_{11} & \frac{a^4}{bm^2} D_{12} \\ \frac{a^4}{bm^2} D_{12} & \frac{a^3}{m^2} D_{22} \end{bmatrix} \quad \text{and} \quad \bar{P} = \frac{\begin{bmatrix} v_1 & -v_3 \\ -v_3 & v_2 \end{bmatrix}}{v_1 v_2 - v_3^2} .$$

The optimal input is full on in one direction and then full on in the opposite direction (bang-bang) with switch times and normalized performance index as shown in Figs. 6.1 and 6.2. The solution shown was calculated for

$$\bar{D} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} .$$

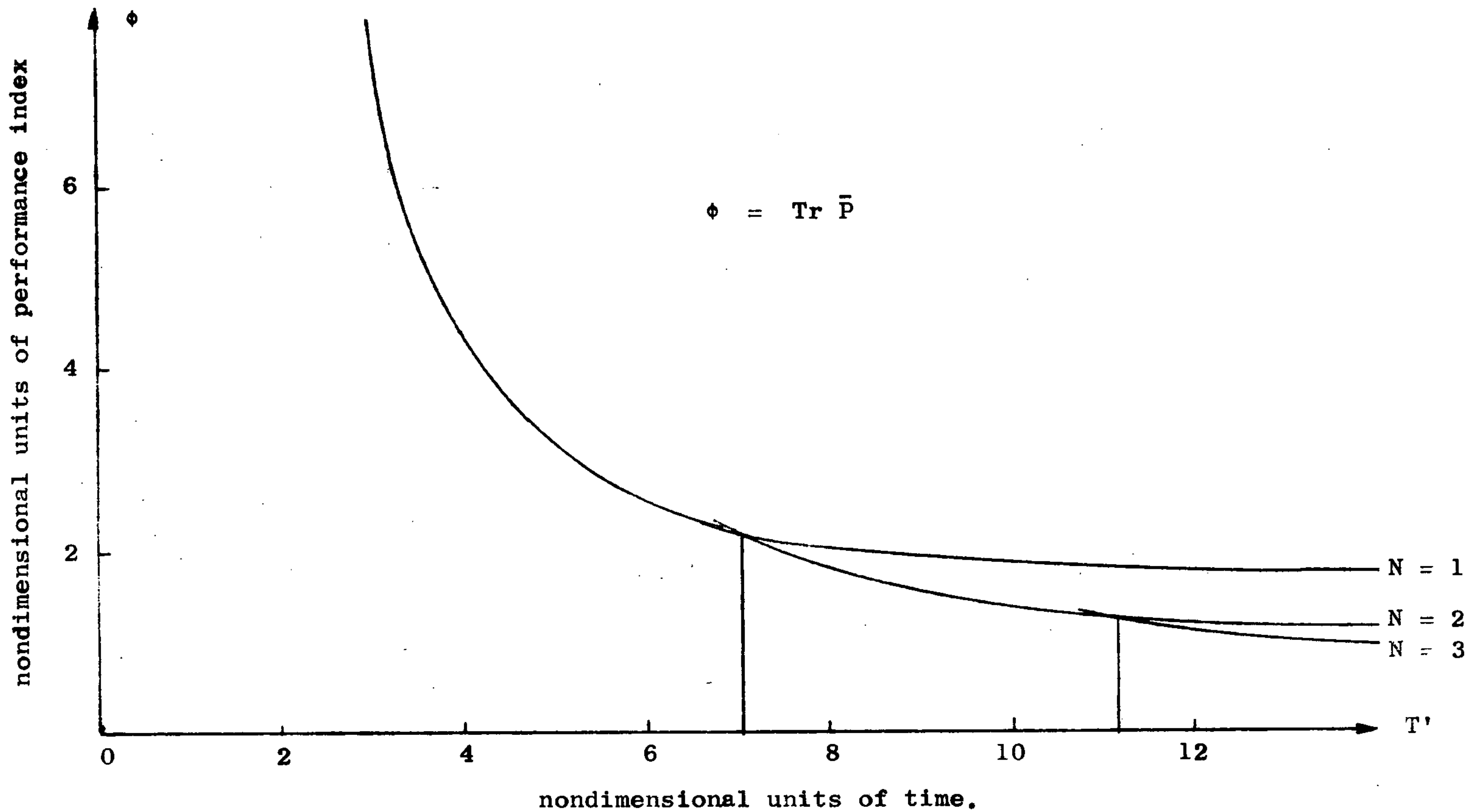


FIG. 6.1. PERFORMANCE INDEX VS LENGTH OF TEST FOR ONE, TWO, AND THREE SWITCHES

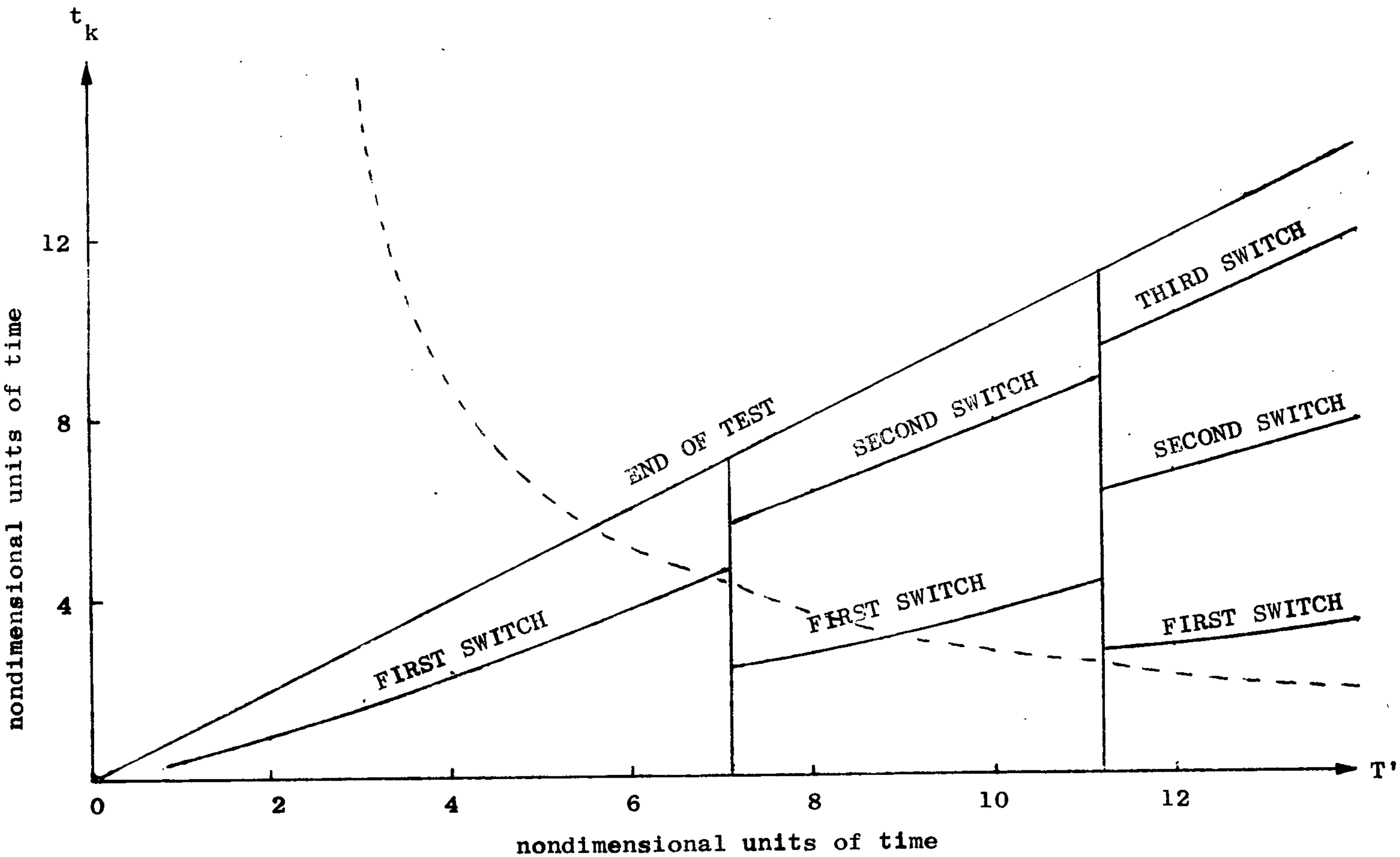


FIG. 6.2. SWITCH TIMES VS LENGTH OF TEST.

Figure 6.1 shows plots of the performance index for one switch (N = 1), through three switches (N = 3), for tests up to 14 time constants. For the no-switch case (N = 0), the performance index Tr $\bar{P}$ asympotically approaches eight and is not shown but is optimal for tests under 0.2 time constants. Figure 6.2 shows the switch times.

If we were to use the suggested criterion of maximizing the trace of I, we would have

$$\phi = \frac{1}{r}\left[\int_0^T \left(\frac{\partial x}{\partial a}\right)^2 dt + \int_0^T \left(\frac{\partial x}{\partial b}\right)^2 dt\right] .$$

It is easy to see from Fig. 6.3 that the optimum input for this criterion is a constant step $u = \pm m$, for any test length. Except for very short tests, the constant input is the worst bang-bang input for minimizing the covariance of the parameters!

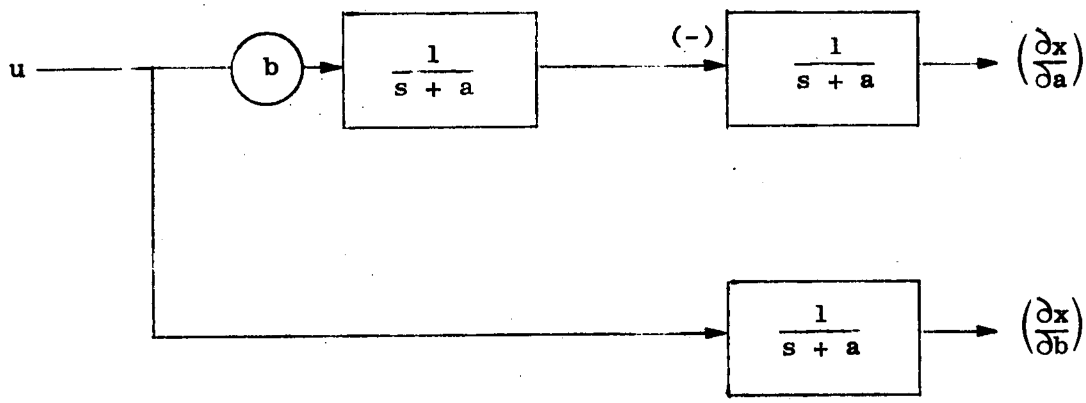


FIG. 6.3. BLOCK DIAGRAM OF STATE AND SENSITIVITY FUNCTIONS.

C. INPUT CRITERIA FROM IDENTIFICATION ALGORITHMS

Now let us look at the identification techniques of the previous Chapter and see if they also include a clue as to an input criterion.

C-1. Quasilinearization

Refer to the summary of the quasilinearization technique in Chapter V. During the second set of iterations, the sensitivity equations are driven by z_N, so that the state, nominal, and sensitivity equations are deterministic. Recall that

$$F = F_N + DH \quad \text{and} \quad G = G_N + \delta G$$

so that

$$\frac{\partial F}{\partial a_i} = \frac{\partial D}{\partial \alpha_i} H + D \frac{\partial H}{\partial \alpha_i} \quad \text{and} \quad \frac{\partial G}{\partial a_i} = \frac{\partial \delta G}{\partial \alpha_i}$$

so that when $F_N = F$, $G_N = G$, $H_N = H$, and $D = 0$, the sensitivity equation in the quasilinearization technique is equal to

$$\dot{\left(\frac{\partial x}{\partial a_i}\right)} = F\left(\frac{\partial x}{\partial a_i}\right) + \left(\frac{\partial F}{\partial a_i}\right) x + \frac{\partial G}{\partial a_i} u, \qquad \frac{\partial x}{\partial a_i}(t_o) = \frac{\partial x_o}{\partial a_i} \quad . \tag{6.13}$$

Also, $H_N = (I-L)H$ so that $(\partial L/\partial \alpha_i)H = (\partial H/\partial a_i)(I-L)$ and when $a_N = a$ we have

$$\left(\frac{\partial \hat{z}}{\partial \alpha_i}\right) = H\left(\frac{\partial x}{\partial a_i}\right) + \frac{\partial H}{\partial a_i} x \ . \tag{6.14}$$

In such a case

$$\hat{\alpha} = I^{-1} \int_{t_o}^{t_f} \left(\frac{\partial \hat{z}}{\partial \alpha}\right)^T R^{-1} v \, dt \tag{6.15}$$

and its mean and covariance are given by

$$\bar{\hat{\alpha}} = 0 \quad \text{and} \quad E\hat{\alpha}\hat{\alpha}^T = I^{-1} \ . \tag{6.16}$$

However, in the algorithm, we have an iterative process that is repeated until $\hat{\alpha} \approx 0$. The value of a_N for which this happens is the identified value of a. The statistics of the resulting $\hat{a}$ are not easily derived. However, we can say that the smaller I^{-1} is, the closer $\hat{a}$ is to the true value of a.

2. Gradient Algorithm

We want to shape the input $u(t)$ so that $\lambda(t_o)$ and $\Gamma(t_o)$ will be large (therefore our gradient will be steeper). In fact, the matrix

$$E \begin{bmatrix} \frac{\partial \lambda(t_o)}{\partial x_o} & \frac{\partial \lambda(t_o)}{\partial a} \\ \frac{\partial \Gamma(t_o)}{\partial x_o} & \frac{\partial \Gamma(t_o)}{\partial a} \end{bmatrix}$$

is the information matrix! This may be seen by referring to Chapter II.D and letting y denotes the augmented vector

$$\begin{pmatrix} x \\ a \end{pmatrix} \text{ and } \psi$$

its adjoint. If $\phi_i \equiv 0$, then $J = J^+(t_o)$ so that

$$\frac{\partial J^{(+)}(t_o)}{\partial y(t_o)} = \psi^T(t_o) .$$

Therefore the information matrix of the state at time t_o, given measurements up through time t_f is given by

$$I(t_o|t_f) = E\frac{\partial \psi(t_o)}{\partial y(t_o)} . \tag{6.17}$$

This is the information matrix we want for identification purposes since we want to identify $x(t_o)$ and $a(t_o)$. If we let

$$\psi = \begin{pmatrix} \lambda \\ \Gamma \end{pmatrix}$$

then we have

$$I(t_o | t_f) = E \begin{bmatrix} \frac{\partial\lambda(t_o)}{\partial x_o} & \frac{\partial\lambda(t_o)}{\partial z} \\ \frac{\partial\Gamma(t_o)}{\partial x_o} & \frac{\partial\Gamma(t_o)}{\partial a} \end{bmatrix} . \tag{6.18}$$

We can now illustrate how the information matrix may be calculated using this indirect approach. The gradient method for the output error criterion gives us the two-point boundary value problem

$$\begin{aligned} \dot{x} &= Fx + Gu, & x(t_o) &= x_o \\ \dot{\lambda} &= H^T R^{-1}(z - Hx) - F^T\lambda, & \lambda(t_f) &= 0 \\ \dot{\Gamma}_i &= x^T \frac{\partial H^T}{\partial a_i} R^{-1}(z-Hx) - \left[x^T \frac{\partial F^T}{\partial a_i} + u^T \frac{\partial G^T}{\partial a_i}\right]\lambda , \end{aligned} \tag{6.19}$$

$$\Gamma_i(t_f) = 0, \qquad i = 1, 2, \ldots q' .$$

The sensitivity equations for $(x^T \lambda^T \Gamma^T)^T$ with respect to $(x_o^T a^T)^T$ are given by

$$\left(\dot{\frac{\partial x}{\partial x_o}}\right) = F\left(\frac{\partial x}{\partial x_o}\right), \qquad \frac{\partial x}{\partial x_o}(t_o) = I$$

$$\left(\dot{\frac{\partial x}{\partial a_i}}\right) = F\left(\frac{\partial x}{\partial a_i}\right) + \frac{\partial F}{\partial a_i} x + \frac{\partial G}{\partial a_i} u, \quad \frac{\partial x}{\partial a_i}(t_o) = 0, \quad i = 1,2,\ldots q' \tag{6.20}$$

and

$$\left(\dot{\frac{\partial \lambda}{\partial x_o}}\right) = -F^T\left(\frac{\partial \lambda}{\partial x_o}\right) - H^T R^{-1} H\left(\frac{\partial x}{\partial x_o}\right), \qquad \frac{\partial \lambda}{\partial x_o}(t_f) = 0 \ ;$$

$$\left(\dot{\frac{\partial \lambda}{\partial a_i}}\right) = -\frac{\partial F^T}{\partial a_i}\lambda - F^T\left(\frac{\partial \lambda}{\partial a_i}\right) + \frac{\partial H^T}{\partial a_i} R^{-1}(z-Hx) - H^T R^{-1}$$

$$\times \left[\frac{\partial H}{\partial a_i} x + H\left(\frac{\partial x}{\partial a_i}\right)\right], \qquad \frac{\partial \lambda}{\partial a_i}(t_f) = 0 \ ;$$

$$\left(\dot{\frac{\partial \Gamma_i}{\partial x_o}}\right) = (z - Hx)^T R^{-1} \frac{\partial H}{\partial a_i}\left(\frac{\partial x}{\partial x_o}\right) - x^T \frac{\partial H}{\partial a_i} R^{-1} H\left(\frac{\partial x}{\partial x_o}\right) - \tag{6.21}$$

$$- \lambda^T \frac{\partial F}{\partial a_i}\left(\frac{\partial x}{\partial x_o}\right) - \left(\frac{\partial F}{\partial a_i} x + \frac{\partial G}{\partial x_i} u\right)^T\left(\frac{\partial \lambda}{\partial x_o}\right), \qquad \frac{\partial \Gamma_i}{\partial x_o}(t_f) = 0 \ ;$$

$$\left(\dot{\frac{\partial \Gamma_i}{\partial a_j}}\right) = (z - Hx)^T R^{-1} \frac{\partial H}{\partial a_i}\left(\frac{\partial x}{\partial a_j}\right) - x^T \frac{\partial H^T}{\partial a_i} R^{-1}\left[\frac{\partial H}{\partial a_j} x + H\left(\frac{\partial x}{\partial a_j}\right)\right] -$$

$$- \left[\frac{\partial F}{\partial a_i}\left(\frac{\partial x}{\partial a_j}\right)^T\right]\lambda - \left(\frac{\partial F}{\partial a_i} x + \frac{\partial G}{\partial a_i} u\right)^T\left(\frac{\partial \lambda}{\partial a_j}\right), \qquad \frac{\partial \Gamma_i}{\partial a_j}(t_f) = 0 \ .$$

(The trivial sensitivities $\frac{\partial a(t)}{\partial a_o}$ and $\frac{\partial a(t)}{\partial x_o}$ imply that $a(t)$ equals a constant, and

$$\frac{\partial a}{\partial x_o} \equiv 0).$$

Taking the expectation of (6.21), we have

$$\left(\overline{\frac{\partial\lambda}{\partial x_o}}\right)^{\bullet} = - F^T\left(\frac{\partial\lambda}{\partial x_o}\right) - H^T R^{-1} H\left(\frac{\partial x}{\partial x_o}\right), \qquad \frac{\partial\lambda}{\partial x_o}(t_f) = 0 \; ;$$

$$\left(\overline{\frac{\partial\lambda}{\partial a_i}}\right)^{\bullet} = - F^T\left(\overline{\frac{\partial\lambda}{\partial a_i}}\right) - H^T R^{-1}\left[\frac{\partial H}{\partial a_i} x + H\left(\frac{\partial x}{\partial a_i}\right)\right],$$

$$\left(\overline{\frac{\partial\lambda}{\partial a_i}}\right)(t_f) = 0, \qquad i = 1, 2, \ldots q' \; ;$$

$$\left(\overline{\frac{\partial\Gamma_i}{\partial x_o}}\right)^{\bullet} = - x^T \frac{\partial H}{\partial a_i} R^{-1} H\left(\frac{\partial x}{\partial x_o}\right) - \left(\frac{\partial F}{\partial a_i} x + \frac{\partial G}{\partial a_i} u\right)^T\left(\frac{\partial\lambda}{\partial x_o}\right), \tag{6.22}$$

$$\left(\overline{\frac{\partial\Gamma_i}{\partial x_o}}\right)(t_f) = 0, \qquad i = 1, 2, \ldots q' \; ;$$

$$\left(\overline{\frac{\partial\Gamma_i}{\partial a_j}}\right)^{\bullet} = - x^T \frac{\partial H}{\partial a_i} R^{-1}\left[\frac{\partial H}{\partial a_j} x + H\left(\frac{\partial x}{\partial a_j}\right)\right] - \left(\frac{\partial F}{\partial a_i} x + \frac{\partial G}{\partial a_i} u\right)^T\left(\overline{\frac{\partial\lambda}{\partial a_j}}\right),$$

$$\left(\overline{\frac{\partial\Gamma_i}{\partial a_j}}\right)(t_f) = 0, \qquad i, j = 1, 2, \ldots q' \, .$$

To find the information matrix (before the test is run), calculate x from (6.19), the sensitivity equations from (6.20), and the mean adjoint sensitivity equations from (6.22). These latter functions at t_o give us the elements of the information matrix. This indirect method involves more computation than the direct method illustrated in the previous section. For the direct method, x and $(\partial x/\partial a_i)$ would have to be calculated but then the elements of the information matrix could be calculated from quadratures of the sensitivity functions. An example of the equivalence of the direct and indirect methods of calculating the information matrix was shown in Chapt. II.D. The example in that section may be viewed as a parameter estimation problem for the final state $x(t_f)$.

3. Nonlinear Filter

It might be interesting to apply one of the nonlinear filter algorithms of Chapt. II.C to our problem. For this discussion, let us assume that a represents the parameters in F and G that are known very poorly: $P_{aa}^{-1}(t_o) \approx 0$; the initial state x_o is known quite well: $P_{xx}(t_o) \approx 0$; we are using a canonical form where H is known and the intensity of the process noise Q is known. By letting $\binom{x}{a}$ be the state in the extended Kalman filter, we have

$$\dot{\hat{x}} = \hat{F}\hat{x} + \hat{G}u + P_{xx}H^T R^{-1}(z - H\hat{x}), \qquad \hat{x}(t_o) = x_o$$

$$\dot{\hat{a}} = P_{ax}H^T R^{-1}(z - H\hat{x}), \qquad \hat{a}(t_o) = a_o$$

$$\dot{P}_{xx} = \hat{F}P_{xx} + \frac{\partial(\widehat{Fx} + \hat{G}u)}{\partial a} P_{xa}^T + P_{xx}\hat{F}^T + $$

$$+ P_{xa}\frac{\partial(\hat{F}\hat{x} + \hat{G}u)^T}{\partial a} + Q - P_{xx}H^T R^{-1} H P_{xx}, \qquad P_{xx}(t_o) = P_o \text{ (small)}$$

$$\dot{P}_{xa} = \hat{F}P_{xa} + \frac{\partial(\widehat{Fx} + \hat{G}u)}{\partial a} P_{aa} - P_{xx}H^T R^{-1} H P_{xa}, \qquad P_{xa}(t_o) = 0$$

$$\dot{P}_{aa} = -P_{xa}^T H^T R^{-1} H P_{xa}, \qquad P_{aa}(t_o) = A \text{ (large)}$$

If $P_{aa}(t_o)$, in addition to $P_{xx}(t_o)$, were small, then this would yield a reasonable estimate of the state and the unknown parameters. However, for the problem as formulated above, we cannot integrate the covariance equations with $P_{xx}(t_o) \approx 0$ and $P_{aa}(t_o) \approx \infty$, or $P_{xx}^{-1}(t_o) \approx \infty$ and $P_{aa}^{-1}(t_o) \approx 0$. (Not to mention the premise that for their derivation, the covariances were assumed small relative to the nonlinearities.)

Since we must have an estimate of a to design an input, we may drop the circumflexes on F and G. Making the definitions $S = P_{xa} P_{aa}^{-1}$ and $I_a = P_{aa}^{-1}$, the last two covariance equations become

$$
\begin{aligned}
\dot{S} &= FS + \frac{\partial(F\hat{x}+Gu)}{\partial a} + (P_{xa}P_{aa}^{-1}P_{xa} - P_{xx})H^T R^{-1} HS, \qquad S(t_o) = 0 \\
\dot{I}_a &= S^T H^T R^{-1} HS, \qquad I_a(t_o) = A^{-1}
\end{aligned}
\tag{6.24}
$$

If we assume that $P_{xx} \approx 0$ and note that $I_{xx}^{-1} = P_{xx} - P_{xa} \cdot P_{aa}^{-1} P_{xa}^T > 0$ so that $0 < P_{xa} P_{aa}^{-1} P_{xa}^T < P_{xx}$, we may drop the last term of the first equation in (6.24). The ith column of S is then the sensitivity of x with respect to a_i and we have the same expression for the information matrix as obtained by previous methods. The interesting point to note in this approach is the interpretation of the matrix of sensitivity functions $S = P_{xa} \cdot P_{aa}^{-1}$.

D. PROCESS NOISE

For a system with process noise, we can use the direct method of calculating the information matrix since we were able to minimize (4.31) with respect to w and obtain a Kalman filter representation of the system. For a sufficiently long test, the identification criterion was to minimize

$$
J = \frac{1}{2} \int_{t_o}^{t_f} (z - H\hat{x})^T R^{-1} (z - H\hat{x}) dt \tag{6.25}
$$

with respect to a, subject to the constraint

$$
\dot{\hat{x}} = F\hat{x} + Gu + K(z - H\hat{x}), \qquad \hat{x}(t_o) = x_o . \tag{6.26}
$$

The i,j th element of the information matrix is given by

$$I_{ij} = E\frac{\partial^2 J}{\partial a_i \partial a_j} = E\int_{t_o}^{t_f}\left(\frac{\partial H}{\partial a_i}\hat{x} + H\hat{x}_i\right)^T R^{-1}\left(\frac{\partial H}{\partial a_j}\hat{x} + H\hat{x}_j\right)dt \qquad (6.27)$$

where x_i denotes $\partial x/\partial a_i$ and is given by

$$\dot{\hat{x}}_i = (F-KH)\hat{x}_i + \left(\frac{\partial F}{\partial a_i} - K\frac{\partial H}{\partial a_i}\right)\hat{x} + \frac{\partial G}{\partial a_i}u + \frac{\partial K}{\partial a_i}(z-H\hat{x}),$$
$$\hat{x}_i(t_o) = \frac{\partial x_o}{\partial a_i}. \qquad (6.28)$$

Since the innovation $\nu = z - H\hat{x}$, is white gaussian noise with intensity matrix R, we may rewrite the state and _i_th sensitivity equations as

$$\dot{\hat{x}} = F\hat{x} + Gu + K\nu, \qquad \hat{x}(t_o) = x_o$$
$$\dot{\hat{x}}_i = (F-KH)\hat{x}_i + \left(\frac{\partial F}{\partial a_i} - K\frac{\partial H}{\partial a_i}\right) + \frac{\partial G}{\partial a_i}u + \frac{\partial K}{\partial a_i}\nu, \qquad \hat{x}_i(t_o) = \frac{\partial x_o}{\partial a_i}. \qquad (6.29)$$

The mean of the state equation $\bar{\hat{x}} \triangleq x$, and the mean of the _i_th sensitivity equation $\bar{\hat{x}}_i \triangleq x_i$ are then given by

$$\dot{x} = Fx + Gu,$$
$$\dot{x}_i = (F-KH)x_i + \left(\frac{\partial F}{\partial a_i} - K\frac{\partial H}{\partial a_i}\right)x + \frac{\partial G}{\partial a_i}u, \qquad x_i(t_o) = \frac{\partial x_o}{\partial a_i}. \qquad (6.30)$$

The covariance matrix $P^{oo} \triangleq E(\hat{x}-x)(\hat{x}-x)^T$, $P^{oi} = E(\hat{x} - x)(\hat{x}_i-x_i)^T$, and $P^{ij} \triangleq E(\hat{x}_i-x_i)(\hat{x}_j-x_j)^T$ are determined from

$$\dot{P}^{oo} = FP^{oo} + P^{oo}F^T + KRK^T, \qquad P^{oo}(t_o) = 0 ;$$

$$\dot{P}^{oi} = FP^{oi} + P^{oo}\left(\frac{\partial F^T}{\partial a_i} - \frac{\partial H^T}{\partial a_i} K^T\right) + P^{oi}(F^T - H^T K^T) + $$

$$+ KR \frac{\partial K^T}{\partial a_i}, \qquad P^{oi}(t_o) = 0 ; \tag{6.31}$$

$$\dot{P}^{ij} = \left(\frac{\partial F}{\partial a_i} - K \frac{\partial H}{\partial a_i}\right)P^{oj} + (F-KH)P^{ij} + P^{oi}\left(\frac{\partial F^T}{\partial a_j} - \frac{\partial H^T}{\partial a_j} K^T\right)$$

$$+ P^{ij}(F^T - H^T K^T) + \frac{\partial K}{\partial a_i} R \frac{\partial K^T}{\partial a_j}, \quad P^{ij}(t_o) = 0 .$$

Performing the expectation in (6.27) we obtain

$$I_{ij} = \int_{t_o}^{t_f} \left(\frac{\partial H}{\partial a_i} x + Hx_i\right)^T R^{-1}\left(\frac{\partial H}{\partial a_j} x + Hx_j\right)dt$$

$$+ \int_{t_o}^{t_f} Tr\left[HP^{ij}H^T + \frac{\partial H}{\partial a_i} P^{oj}H^T + HP^{io} \frac{\partial H^T}{\partial a_j} + \right. \tag{6.32}$$

$$\left. + \frac{\partial H}{\partial a_i} P^{oo} \frac{\partial H^T}{\partial a_j}\right] R^{-1} dt .$$

The positive definite covariances in the information matrix imply that process noise may increase the accuracy of our estimate. However, we should note that the new sensitivity equations (6.30) that act as constraint equations in our optimization, are also modified by the process noise. For a simple example shown in the next Chapter, process noise tends to decrease the effectiveness of the input, so that the net effect is a decrease of estimation accuracy with process noise.

Chapter VII

SOLUTION FOR OPTIMAL INPUTS

A. INTRODUCTION

We have seen that a reasonable criterion for judging inputs to identify q parameters in a linear system is some measure of the information matrix of the parameters we wish to identify. To evaluate this criterion, we must solve n linear system equations which drive $n \cdot q$ sensitivity equations, which, in turn are used to generate $\frac{1}{2}q(q + 1)$ elements of this information matrix.

Optimization of this criterion can be formulated as a calculus-of-variations problem (Mayer formulation) to minimize $\phi[y(t_f)] = \mathrm{Tr}DI^{-1}(t_f)$ subject to the constraints

$$\dot{y} = f(y) + Bu, \qquad y(t_o) = y_o, \qquad |u| \leq m \tag{7.1}$$

where y represents the state, sensitivity functions, and elements of the information matrix. The dimension of y is then $(n + \frac{1}{2}q)(q + 1)$. For the general case (with process noise), the constraints in (7.1) are given by

$$\dot{x} = Fx + Gu, \qquad x(t_o) = x_o \; ; \tag{7.2}$$

$$\dot{x}_i = (F\text{-}KH)x_i + \left(\frac{\partial F}{\partial a_i} - K\frac{\partial H}{\partial a_i}\right)x + \frac{\partial G}{\partial a_i}, \quad x_i(t_o) = \frac{\partial x_o}{\partial a_i}, \quad i = 1,2,\ldots q; \tag{7.3}$$

$$\dot{I}_{ij} = \left(\frac{\partial H}{\partial a_i}x + Hx_i\right)^T R^{-1}\left(\frac{\partial H}{\partial a_j}x + Hx_j\right) + C_{ij}, \quad I_{ij}(t_o) = A_{ij}^{-1},$$

$$i = 1, 2, \ldots q, \qquad j = i, i + 1, \ldots q \tag{7.4}$$

where C_{ij} represents the second integrand in (6.32). For the case without process noise, C_{ij} and K are equal to zero.

From the linearity of u in the constraint equation (7.1), with its absence from the performance index, and its amplitude constraint, we have, from Pontryagin's maximum principle, that the optimal input is bang-bang with amplitude m. All that remains is to find the switch times that optimize the performance index.

If we let λ, λ_i (vectors) and λ_{ij} be adjoint variables corresponding to x, x_i, and I_{ij} respectively, we can form the Hamiltonian

$$\mathcal{H} = \lambda^T(Fx + Gu) + \sum_{i=1}^{q} \lambda_i^T\left[(F-KH)x_i + \left(\frac{\partial F}{\partial a_i} - K\frac{\partial H}{\partial a_i}\right)x + \frac{\partial G}{\partial a_i}u\right] + \sum_{i=1}^{q}\sum_{j=i}^{q}\lambda_{ij}\left[\left(\frac{\partial H}{\partial a_i}x + Hx_i\right)^T R^{-1}\left(\frac{\partial H}{\partial a_j}x + Hx_j\right) + C_{ij}\right]. \tag{7.5}$$

The Euler-Lagrange equations for the conjugate variables are

$$\dot{\lambda}^T = -\lambda^T F - \sum_{i=1}^{q}\lambda_i^T\left(\frac{\partial F}{\partial a_i} - K\frac{\partial H}{\partial a_i}\right) - \sum_{i=1}^{q}\sum_{j=i}^{q}\lambda_{ij} \times \left[\left(\frac{\partial H}{\partial a_i}x + Hx_i\right)^T R^{-1}\frac{\partial H}{\partial a_j} + \left(\frac{\partial H}{\partial a_j}x + Hx_j\right)^T R^{-1}\frac{\partial H}{\partial a_i}\right],$$
$$\lambda(t_f) = 0 ; \tag{7.6}$$

$$\dot{\lambda}_i^T = -\lambda_i^T(F-KH) - \sum_{j=i}^{q}\lambda_{ij}\left(\frac{\partial H}{\partial a_j}x + Hx_j\right)^T R^{-1}H - \sum_{j=1}^{i}\lambda_{ji}\left(\frac{\partial H}{\partial a_j}x + Hx_j\right)^T R^{-1}H, \quad \lambda_i(t_f) = 0, \quad i = 1,2,\dots q . \tag{7.7}$$

$$\dot{\lambda}_{ij} = 0$$

$$\lambda_{ij} = \text{constant} = \frac{\partial\phi}{\partial I_{ij}(t_f)}, \quad i = 1,2, \ldots q \qquad (7.8)$$

$$j = i, \; i+1, \ldots q .$$

To minimize the Hamiltonian, the ith component of the input vector u must satisfy the equation

$$u_i = -m_i \operatorname{sgn} S_i \qquad (7.9)$$

where the switching functions S_i are given by

$$S_i = \frac{\partial H}{\partial u_i} = \lambda^T G_i + \sum_{j=1}^{q} \lambda_j^T \frac{\partial G_i}{\partial a_j} \qquad (7.10)$$

where G_i denotes the ith column of G. In this formulation we must find an input u that satisfies (7.2) to (7.4) and (7.6) to (7.10). One algorithm for this is

(1) Choose an initial switching sequence for the input;

(2) Integrate (7.2) to (7.4) forward with the given initial conditions;

(3) Calculate the constants λ_{ij} from (7.8) and integrate (7.6) and (7.7) backward.

(4) Calculate the switching function(s) by use of (7.10). If the optimality condition, (7.9) is satisfied, then terminate the algorithm, otherwise continue.

(5) Use some criterion to modify the switch sequences so that (hopefully) the next iteration will be closer. One method suggested by Ichikawa and Tamura [IC-1], is to locate the minimums and maximums of the switching functions and expand the corresponding switch intervals out from these points. Create new switch intervals at minimums and maximums as necessary.

(6) With the new switching sequences, go back to step (2) above.

An analysis of the computation required shows that we must integrate $n + q \cdot n$ first order differential equations and $\frac{1}{2}q(q + 1)$ quadratures forward and $n + q \cdot n$ equations backwards for each iteration of this algorithm. Hence, there are $(q + 1)(2n + \frac{1}{2}q)$ integrations per iteration. This number is independent of the number of switches.

A-1. Steady-State Sine Input

For long tests the optimal input is often a bang-bang input with almost equal switch time intervals. In this case, a good approximation to the minimum value of the performance index and the optimum switch times can be obtained by approximating the optimal square wave by its first (and possibly higher order) Fourier component(s). Since we are assuming a long test time, we may use the steady-state amplitude ratio and phase shift calculated from the transfer function. We then have only p angular frequencies ω_i, $i = 1,2,\dots p$ to optimize. If two input frequencies are the same, then we would also have to optimize with respect to their phase.

B. OPTIMAL INPUT ALGORITHM

Since we know that the optimal input is bang-bang, we can optimize with respect to the switch times, t_1, $t_2,\dots t_N$. To insure a global minimum, the optimal value of the performance index may be plotted versus the length of the test for $N = 0, 1, 2 \dots$. For example, see Fig. 7.5 of section 7.D where for $T = 8$, there is a minimum for $N = 0, 1, 2, 3$. The minimum for $N = 2$ is the global minimum. The algorithm of the previous section could converge to any of the local minima. It could not be used in the systematic method outlined above since it creates and annihilates switch times as necessary.

The algorithm that seems most promising for determining the optimum switch times is the conjugate gradient algorithm. Using this method, the minimum of a quadratic function of N parameters is found in N iterations. The first iteration involves searching in the steepest descent direction until a minimum along that direction is found. On subsequent iterations, the search is made in a conjugate direction.

The implementation of this algorithm to our problem is shown in the flow diagram of Fig. 7.1 and follows Pierre [PI-1].

The following features concerned with the one dimensional search have been incorporated into the algorithm:

(1) If the one-dimensional search finds a minimum at a distance greater than three times the value of the initial step size, A_1 then a steepest descent search is continued;

(2) The initial A_1 is taken as 1/5 the initial time interval multiplied by the number of switch times. For a set of N iterations the same value of A_1 is used;

(3) For a new set of iterations, the value of A_1 is set equal to 1/5 the average search distances for the previous N iterations;

(4) A quadratic or cubic fit is used to find the minimum in the one dimensional search.

2. One Dimensional Search

The one dimensional search algorithm is shown in Fig. 7.2. Let r be the direction vector in the t_1 through t_N space given by

$$r = Hg \tag{7.11}$$

where H is a matrix given by the conjugate gradient algorithm and g is the gradient of the performance index. A change in the kth switch time in the -r direction is given by

$$\delta t_k = - \frac{r_k A}{R} \tag{7.12}$$

where

$$R = \sum_{k=1}^{N} |r_k| \quad .$$

Since

$$\left|\frac{r_k}{R}\right| \leq 1 ,$$

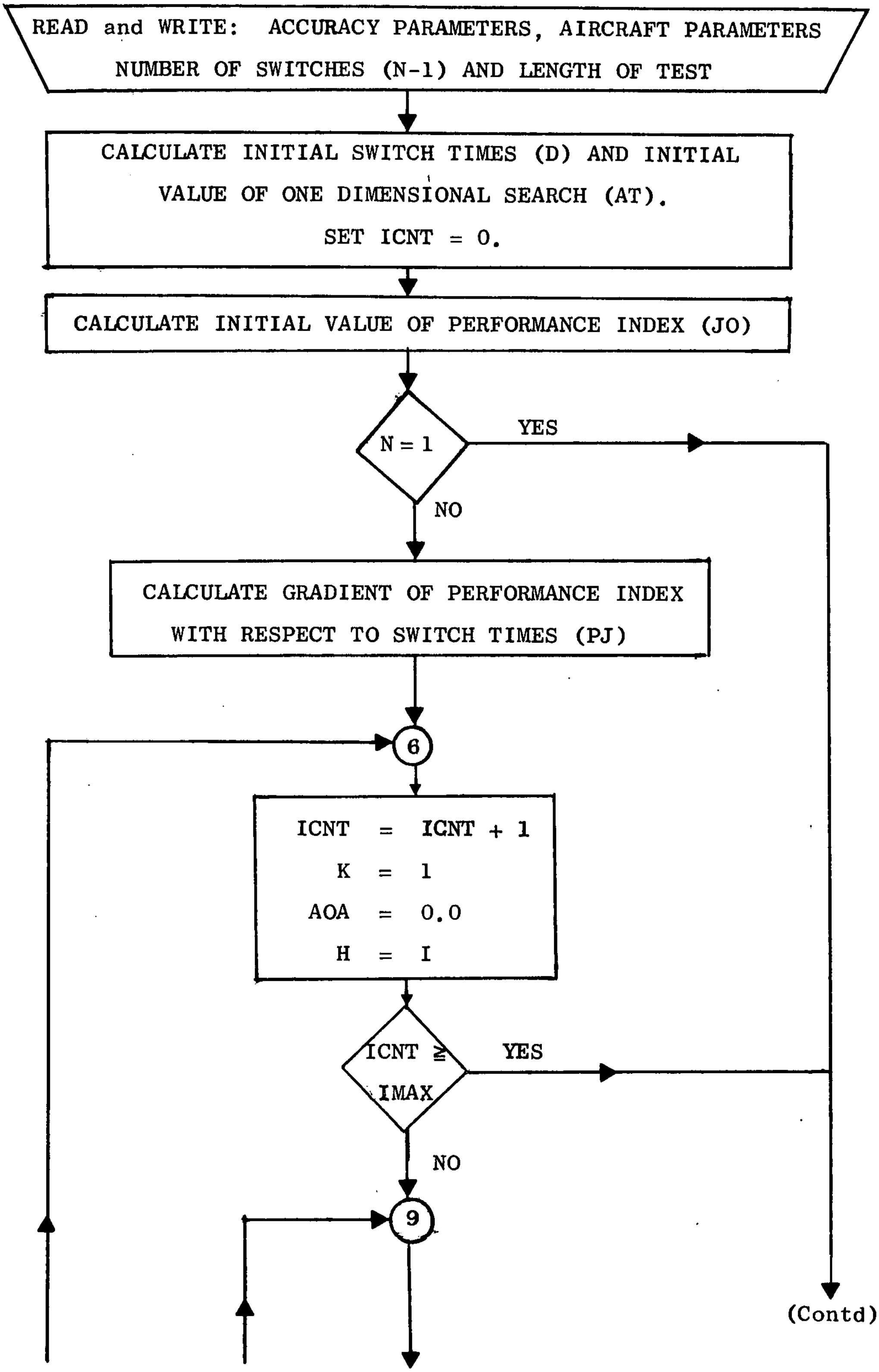


FIG. 7.1 FLOW DIAGRAM OF CONJUGATE GRADIENT ALGORITHM

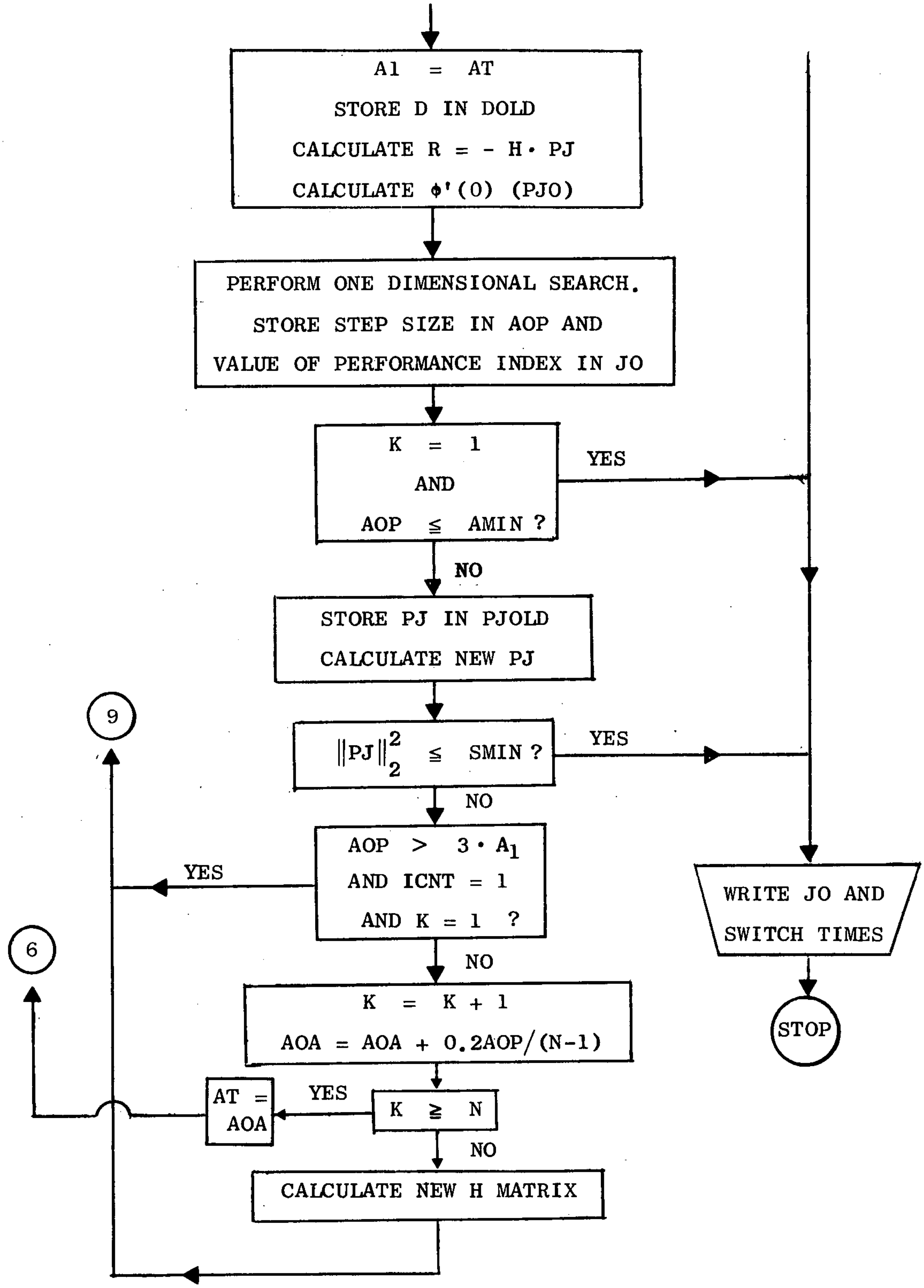


FIG. 7.1 (Conclusion)

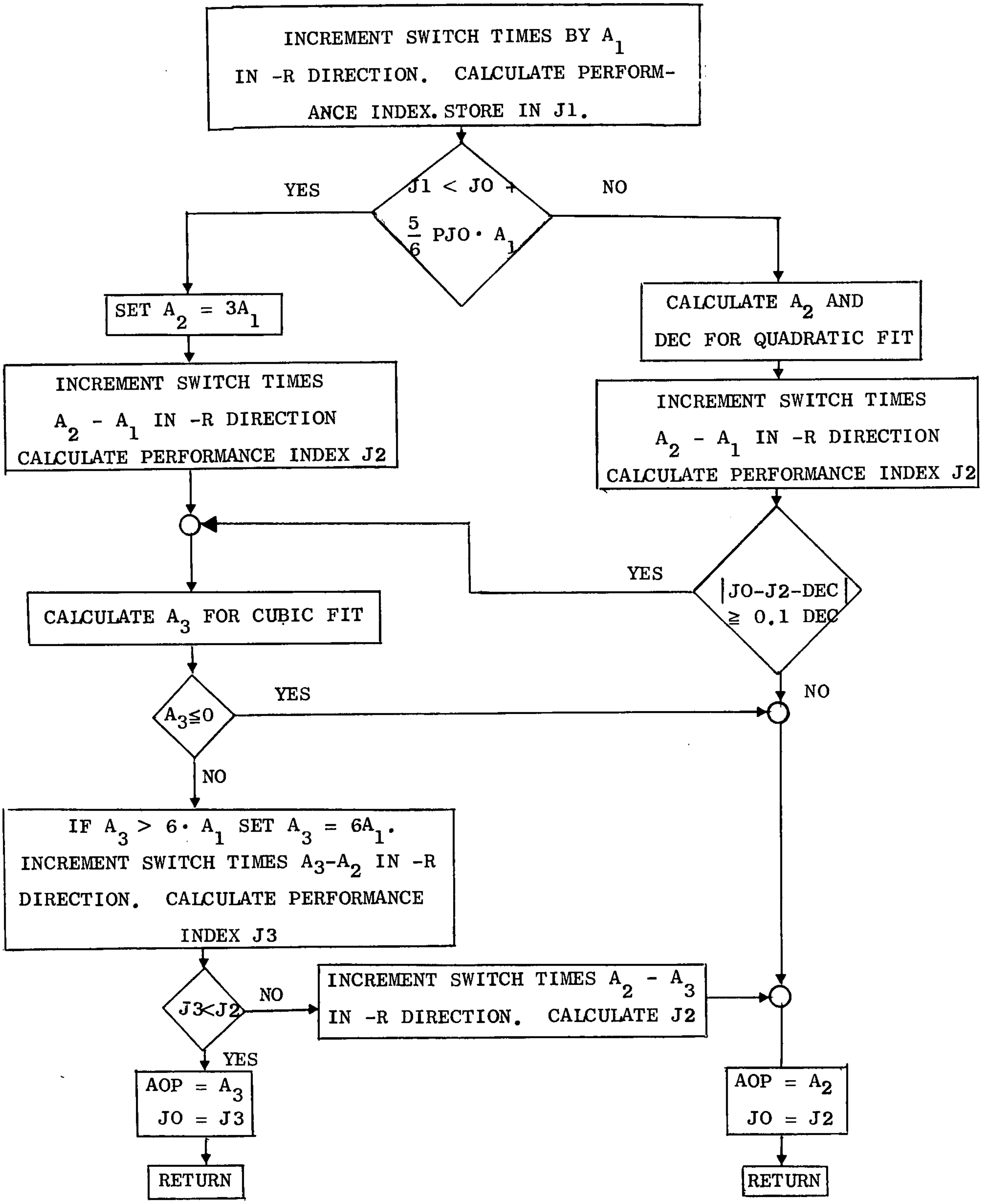


FIG. 7.2 FLOW DIAGRAM OF ONE DIMENSIONAL SEARCH ALGORITHM

no switch time changes by more than $\pm A$, and the sum of the absolute values of the changes in switch times is

$$\sum_{k=1}^{N} |\delta t_k| = \sum_{k=1}^{N} |r_k| \frac{A}{R} = A \,. \tag{7.13}$$

To first-order a change in the performance index in the $-r$ direction is thus

$$\delta\phi = -(g_1 \cdot r_1 + g_2 r_2 + \cdots g_N \cdot r_N) \frac{A}{R} = -r^T g \frac{A}{R} \,. \tag{7.14}$$

In the one dimensional search portion of this algorithm, we desire to find the value of A which minimizes the performance index in the $-r$ direction. ϕ may be considered a function of A with $\phi(0)$ given and

$$\phi'(0) \triangleq \left.\frac{\partial\phi(A)}{\partial A}\right|_{A=0} = -\frac{r^T g}{R} \,. \tag{7.15}$$

A step of A_1 is taken in the $-r$ direction and the performance index $\phi(A_1)$ is calculated. Since we have normalized our gradient, A_1 may be chosen as the maximum total expected change in the switch times, say, 1/5 the interval between switch times multiplied by the number of switch times.

If we are sufficiently close to the optimum (so that a quadratic fit is a good approximation), then ϕ may be written in the form

$$\phi = a + bA + cA^2 \tag{7.16}$$

where the constants a, b, and c are given by

$$a = \phi(0)$$

$$b = \phi'(0)$$

$$c = \frac{\phi(A_1) - \phi(0) - \phi'(0)A_1}{A_1^2} \,.$$

The minimum occurs at

$$A_2 = -\frac{b}{2c} = \frac{-\frac{1}{2}\phi'(0)A_1^2}{\phi(A_1) - \phi(0) - \phi'(0)A_1} \quad . \tag{7.18}$$

Before using (7.18) as the next step size in the one dimensional search, we should check to see if $\phi(A_1)$ is less than or only slightly greater than $\phi(0) + \phi'(0)A_1$. Let us set an upper limit on A_2 of $3A_1$ whenever

$$\phi(A_1) \leq \phi(0) + \phi'(0)A_1 - \frac{1}{6}\phi'(0)A_1 \tag{7.19}$$

and proceed with a cubic fit. Whenever A_2 is less than $3A_1$, the predicted decrease is

$$\text{dec} = \phi(0) - \phi_{\text{pred}} = \frac{\frac{1}{4}[\phi'(0)A_1]^2}{\phi(A_1) - \phi(0) - \phi'(0)A_1} \tag{7.20}$$

If the actual decrease $\phi(0) - \phi(A_2)$ is not close to the predicted decrease, we should go to a cubic fit. Otherwise, the quadratic approximation is sufficient for this one dimensional search.

For a cubic fit we approximate ϕ by

$$\phi = a + bA + cA^2 + dA^3 \tag{7.21}$$

where the constants a, b, c, and d are given by

$$a = \phi(0)$$

$$b = \phi'(0)$$

$$c = e_1 - d \cdot A$$

$$d = (e_1 - e_2)/(A_1 - A_2) \tag{7.22}$$

$$e_1 = \frac{1}{A_1^2}[\phi(A_1) - \phi(0) - \phi'(0)A_1]$$

$$e_2 = \frac{1}{A_2^2}[\phi(A_2) - \phi(0) - \phi'(0)A_2] \; . \tag{7.22 Cont.}$$

The minimum occurs at

$$A_3 = \frac{-c + \sqrt{c^2 - 3bd}}{3d} \; . \tag{7.23}$$

3. Calculation of Performance Index and Gradient

To compute the performance index ϕ for a set of switch times t_1 through t_N requires integrating equations (7.2) to (7.4). The partial derivatives of the performance index with respect to the switch times are functions of $I_{ij}(t_f)$ and $\partial I_{ij}(t_f)/\partial t_k$, namely

$$\frac{\partial \phi}{\partial t_k} = - \operatorname{Tr} I^{-1} \frac{\partial I}{\partial t_k} I^{-1} \; . \tag{7.24}$$

These are given by

$$\frac{\partial x}{\partial t_k} = 0, \quad \frac{\partial x_i}{\partial t_k} = 0, \quad \frac{\partial I_{ij}}{\partial t_k} = 0 \; , \tag{7.25}$$

for $t < t_k$ and

$$\frac{\partial x}{\partial t_k} = \dot{x}\Big|_{t=t_k^-} - \dot{x}\Big|_{t=t_k^+} = G\left[u(t_k^-) - u(t_k^+)\right]$$

$$\frac{\partial x_i}{\partial t_k} = \dot{x}_i\Big|_{t=t_k^-} - \dot{x}_i\Big|_{t=t_k^+} = \frac{\partial G}{\partial a_i}\left[u(t_k^-) - u(t_k^+)\right] \tag{7.26}$$

F

$$\frac{\partial I_{ij}}{\partial t_k} = \dot{I}_{ij}\bigg|_{t=t_k^-} - \dot{I}_{ij}\bigg|_{t=t_k^+} = 0 \qquad (7.26)\ \text{Cont.}$$

for $t = t_k$. For $t > t_k$, the partial derivatives are found by solving

$$\left(\frac{\partial \dot{x}}{\partial t_k}\right) = F\left(\frac{\partial x}{\partial t_k}\right)$$

$$\left(\frac{\partial \dot{x}_i}{\partial t_k}\right) = (F - KH)\left(\frac{\partial x_i}{\partial t_k}\right) + \left(\frac{\partial F}{\partial a_i} - K\frac{\partial H}{\partial a_i}\right)\left(\frac{\partial x}{\partial t_k}\right) \qquad (7.27)$$

$$\left(\frac{\partial \dot{I}_{ij}}{\partial t_k}\right) = \left(\frac{\partial H}{\partial a_i}\frac{\partial x}{\partial t_k} + H\frac{\partial x_i}{\partial t_k}\right)^T R^{-1}\left(\frac{\partial H}{\partial a_j}x + Hx_j\right) + \left(\frac{\partial H}{\partial a_i}x + Hx_i\right)^T R^{-1}\left(\frac{\partial H}{\partial a_j}\frac{\partial x}{\partial t_k} + H\frac{\partial x_j}{\partial t_k}\right)$$

with initial conditions at $t = t_k$ given by (7.26).

To compute the performance index involves the integration of $n + q \cdot n$ first order differential equations and $\frac{1}{2}q(q + 1)$ quadratures. To compute the gradient of the performance index requires integrating (7.27) with initial conditions given by (7.26). Since (7.27) requires x and x_i, (7.2) and (7.3) must also be integrated (unless their values have been stored). Although this involves $N\ (n + nq + n + nq)$ differential equations and $N\frac{1}{2}q(q + 1)$ quadratures, they are not integrated the entire length of the test. The computation involved is equivalent to $\frac{1}{2}N(q + 1)\ (2n + \frac{1}{2}q)$ integrations the entire length of the test. For $N > 2$ this algorithm involves more computation per iteration than the algorithm suggested in the previous section. However, it is still used for the reason given at the beginning of this section.

C. EXAMPLE 1: ROCKET SLED TEST*

An accelerometer is modelled by the equation

$$y = (1 + c_1)u + c_2u^2 \tag{7.28}$$

where y is the output of the accelerometer, and u is the acceleration. In order to evaluate the constants c_1 and c_2, the accelerometer is mounted on a rocket sled. The sled has a maximum acceleration m_1, and can be water-braked with a maximum deceleration m_2. If we assume that the accelerometer measurement is corrupted by white noise v, with zero mean and spectral density r, then the measurement is given by

$$z = (1 + c_1)u + c_2u^2 + v . \tag{7.29}$$

The identification performance index becomes

$$J = \frac{1}{2r}\int_0^T \left\{z - \left[(1 + c_1)u + c_2u^2\right]\right\}^2 dt . \tag{7.30}$$

Since J is quadratic in c_1 and c_2, the likelihood equation $\partial J/\partial c = 0$ is linear in c_1 and c_2:

$$\begin{aligned} \frac{\partial J}{\partial c_1} &= \frac{1}{r}\int_0^T \left[z - (1 + c_1)u - c_2u^2\right](-)u\, dt = 0 \\ \frac{\partial J}{\partial c_2} &= \frac{1}{r}\int_0^T \left[z - (1 + c_1)u - c_2u^2\right](-)u^2 dt = 0; \end{aligned} \tag{7.31}$$

or

$$\int_0^T u^2 dt \cdot c_1 + \int_0^T u^3 dt \cdot c_2 = \int_0^T (z - u)u\, dt , \tag{7.32}$$

---

* This Example suggested by Paul Kaminski [KAM-1]

$$\int_0^T u^3 dt \cdot c_1 + \int_0^T u^4 dt \cdot c_2 = \int_0^T (z - u)u^2 dt \qquad (7.32) \text{ Cont.}$$

Estimates of c_1 and c_2 are given by

$$\begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} \int_0^T u^2 dt & \int_0^T u^3 dt \\ \int_0^T u^3 dt & \int_0^T u^4 dt \end{bmatrix}^{-1} \begin{bmatrix} \int_0^T (z - u)u\, dt \\ \int_0^T (z - u)u^2 dt \end{bmatrix} . \qquad (7.33)$$

The information matrix for c_1 and c_2 is

$$I = E \frac{\partial^2 J}{\partial c^2} = \frac{1}{r} \begin{bmatrix} \int_0^T u^2 dt & \int_0^T u^3 dt \\ \int_0^T u^3 dt & \int_0^T u^4 dt \end{bmatrix} . \qquad (7.34)$$

The lower bound for the covariance matrix of c_1 and c_2 becomes

$$P = r \frac{\begin{bmatrix} x_4(T) & -x_3(T) \\ -x_3(T) & x_2(T) \end{bmatrix}}{x_2(T)x_4(T) - x_3^2(T)} \qquad (7.35)$$

where

$$x_i(t) \triangleq \int_0^t u^i \, dt . \qquad (7.36)$$

The terminal boundary condition $x_1(T) = 0$ is required for the sled to be at rest at the end of the test.

The identification performance index is used to find an estimate of the parameters, and the input performance index is a measure of the accuracy of the identification. The measure of the covariance matrix we desire to minimize depends upon the purpose of our identification. Since the output is of the form $z = (1 + c_1)u + c_2u^2 + v$, an estimate of the acceleration $\hat{u}$ is made from the following (assuming $|c_1| \ll 1$, and $|c_2| \ll 1/u$)

$$\hat{u} = \frac{-(1+\hat{c}_1) + \sqrt{(1+\hat{c}_1)^2 + 4\hat{c}_2 z}}{2\hat{c}_2} \cong z - \hat{c}_1 z - \hat{c}_2 z^2 \tag{7.37}$$

so that $\hat{u}$ is approximated by

$$\hat{u} = u + (c_1 - \hat{c}_1)u + (c_2 - \hat{c}_2)u^2 + v \tag{7.38}$$

and the error in the corrected accelerometer output is

$$\delta\hat{u} = \delta c_1 \cdot u + \delta c_2 u^2 + v . \tag{7.39}$$

If the instrument is to be used at an acceleration level, a, then we would like to minimize the error at that acceleration so that our input performance index is given by

$$\phi = E(\delta u)^2 = \mathrm{tr} \begin{bmatrix} a^2 & a^3 \\ a^3 & a^4 \end{bmatrix} \begin{bmatrix} E\delta c_1^2 & E\delta c_1 \delta c_2 \\ E\delta c_1 \delta c_2 & E\delta c_2^2 \end{bmatrix} = \mathrm{tr}\ DP_c . \tag{7.40}$$

Notice that for this problem, the weighting matrix D is independent of the unknown constants. Our problem then is to minimize

$$\phi = \frac{a^2 x_4(T) - 2a^3 x_3(T) + a^4 x_2(T)}{x_2(T)x_4(T) - x_3^2(T)} \tag{7.41}$$

subject to the constraints

$$\begin{aligned} \dot{x}_1 &= u , & x_1(0) &= 0 , & x_1(T) &= 0 \\ \dot{x}_2 &= u^2 , & x_2(0) &= 0 & & \\ \dot{x}_3 &= u^3 , & x_3(0) &= 0 & & \\ \dot{x}_4 &= u^4 , & x_4(0) &= 0 & & \end{aligned}$$

and

$$-m_2 \leqq u \leqq m_1 . \tag{7.43}$$

The Hamiltonian for this problem is a quartic in u:

$$\mathcal{H} = \lambda_1 u + \lambda_2 u^2 + \lambda_3 u^3 + \lambda_4 u^4 \tag{7.44}$$

where

$$\dot{\lambda}_i = 0 \quad \text{or} \quad \lambda_i = \text{constant}, \quad i = 1, 2, 3, 4 \tag{7.45}$$

where (assuming $a = 1$),

$$\lambda_2 = - \frac{x_3^2(T) + x_4^2(T)}{[x_2(T)x_4(T) - x_3^2(T)]^2} = \text{a negative number}$$

$$\lambda_3 = \frac{2x_3(T)[x_2(T) + x_4(t)]}{[x_2(T)x_4(T) - x_3^2(T)]^2} \tag{7.46}$$

$$\lambda_4 = - \frac{x_2^2(T) + x_3^2(T)}{[x_2(T)x_4(T) - x_3^2(T)]^2} = \text{a negative number.}$$

If the boundary condition $x_1(T) = 0$ is to be satisfied, then only the four possibilities shown in Figs. 7.3a to 7.3d are possible.

The possibility of one and only one intermediate (constant) value m_o is a consequence of the Hamiltonian being a quartic function of u (and the λ coefficients being constants).

The first three possibilities shown in Figs. 7.3a through 7.3c are considered special cases of the fourth possibility. If u equals m_1 for t_1 seconds, $-m_2$ for t_2 seconds, and m_o for t_o seconds, then

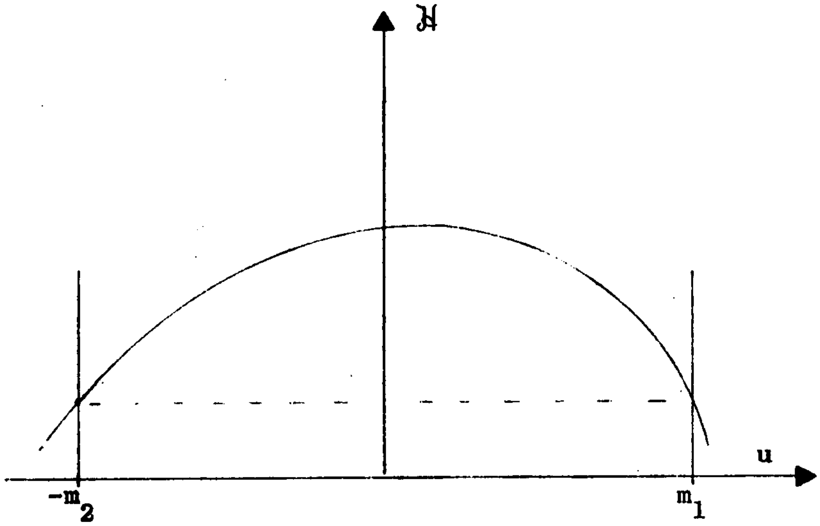


Fig. 7.3a: Case 1: u_{opt} either m_1 or $-m_2$

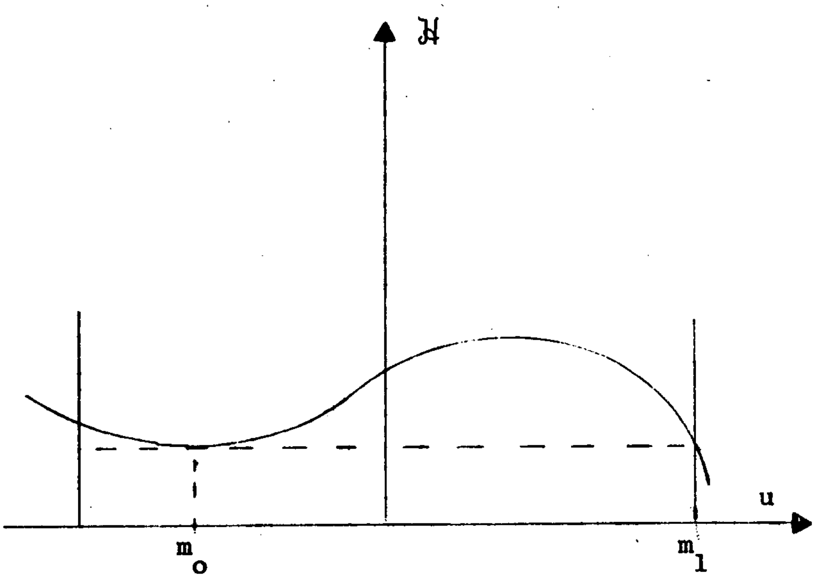


Fig. 7.3b Case 2: u_{opt} either m_1 or m_o

FIGS. 7.3a and 7.3b HAMILTONIAN VS CONTROL

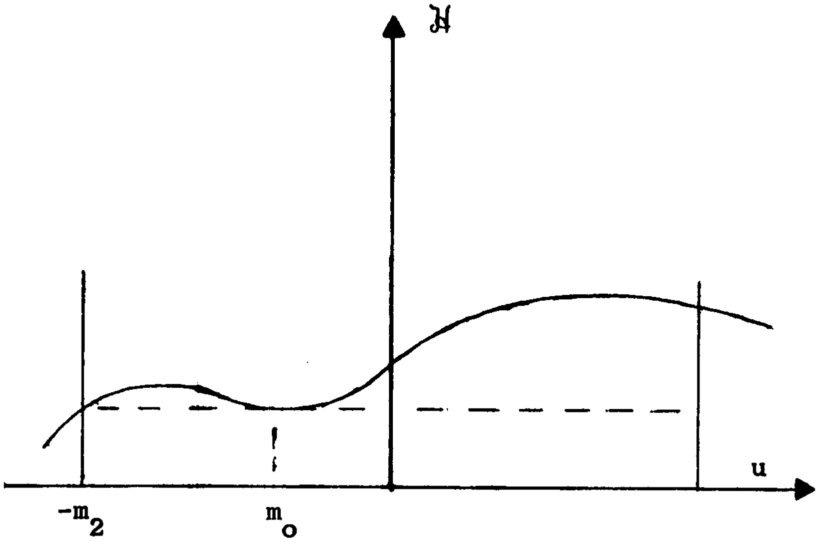


Fig. 7.3c. Case 3: u_{opt} either m_o or $-m_2$

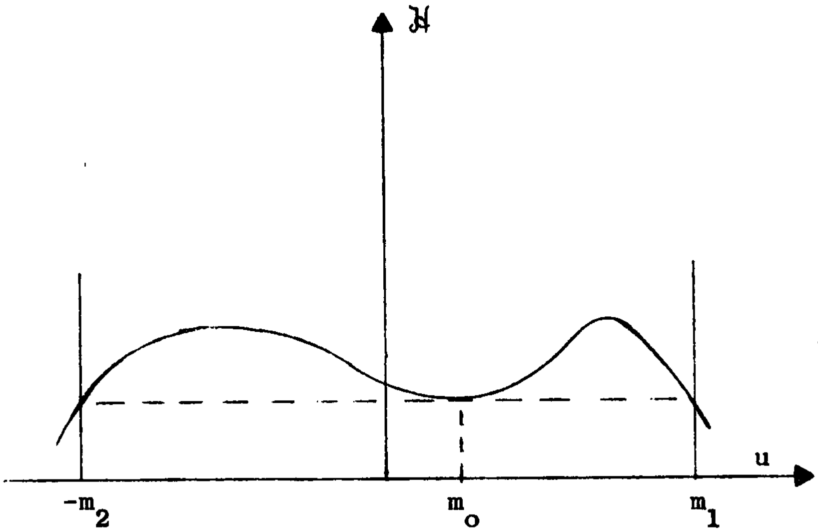


Fig. 7.3d. Case 4: u_{opt} either m_1, m_o or $-m_2$

FIGS. 7.3c and 7.3d HAMILTONIAN VS CONTROL (Cont)

$$
\begin{aligned}
T &= t_1 + t_2 + t_o \\
x_1 &= m_1 t_1 - m_2 t_2 + m_o t_o = 0 \\
x_2 &= m_1^2 t_1 + m_2^2 t_2 + m_o^2 t_o \\
x_3 &= m_1^3 t_1 - m_2^3 t_2 + m_o^3 t_o \\
x_4 &= m_1^4 t_1 + m_2^4 t_2 + m_o^4 t_o \, .
\end{aligned}
\tag{7.47}
$$

The first two equations in (7.47) imply that

$$
\begin{aligned}
t_1 &= \frac{m_2}{m_1 + m_2}(T - t_o) - \frac{m_o}{m_1 + m_2} t_o \geq 0 \\
t_2 &= \frac{m_1}{m_1 + m_2}(T - t_o) + \frac{m_o}{m_1 + m_2} t_o \geq 0 \, .
\end{aligned}
\tag{7.48}
$$

The above inequalities tell us that

$$
\begin{aligned}
t_o &\leq \frac{m_2}{m_2 + m_o} T \\
t_o &\leq \frac{m_1}{m_1 - m_o} T \, .
\end{aligned}
\tag{7.49}
$$

The allowable region for m_o and t_o is then given by

$$
\begin{aligned}
&-m_2 \leq m_o \leq m_1 \\
&0 \leq t_o \leq \frac{m_2}{m_2 + m_o} T \quad \text{for} \quad m_o \geq 0 \\
&0 \leq t_o \leq \frac{m_1}{m_1 - m_o} T \quad \text{for} \quad m_o \leq 0
\end{aligned}
\tag{7.50}
$$

and is shown in Fig. 7.4.

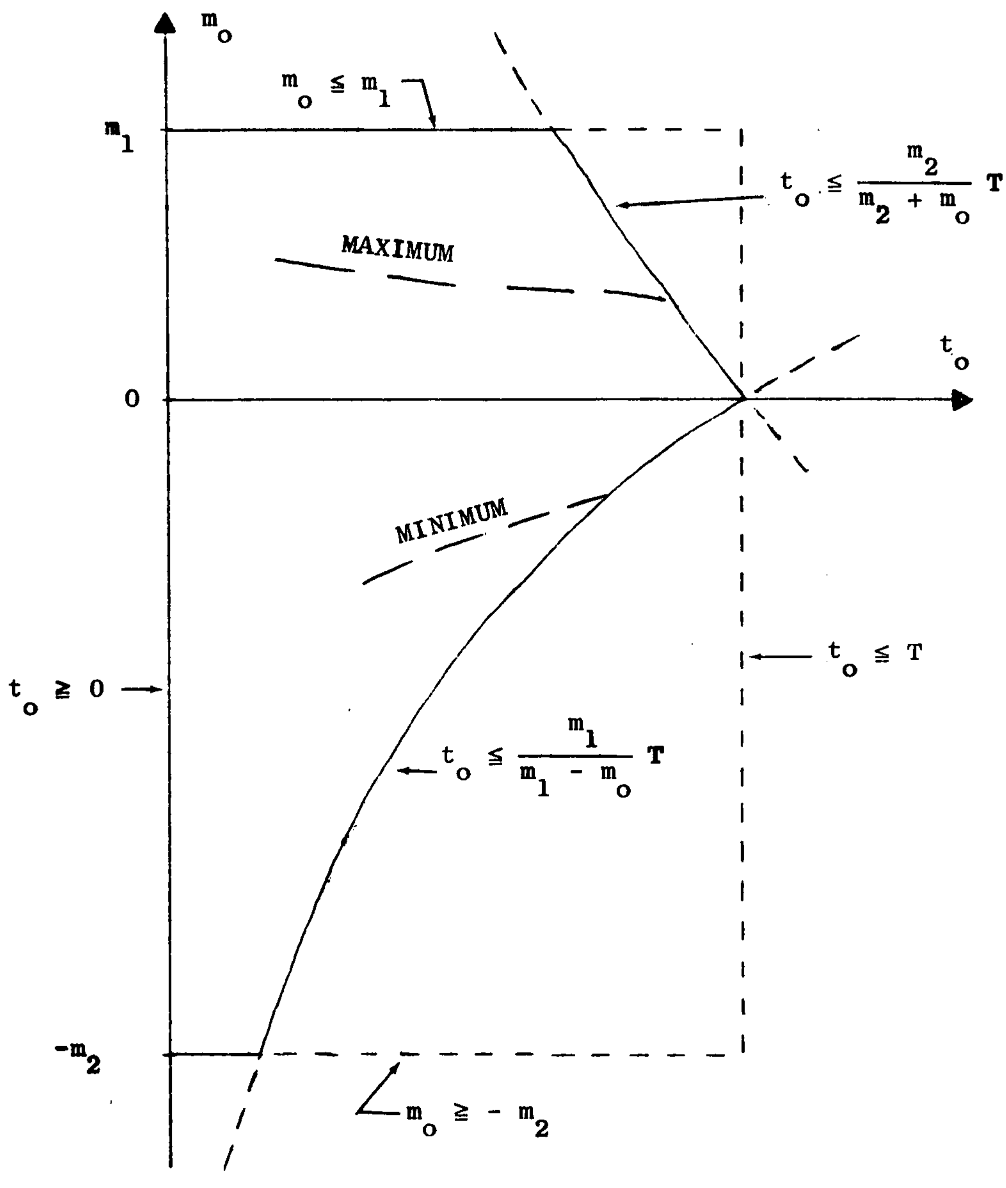


FIG. 7.4 ALLOWABLE REGION FOR m_o AND t_o .

In summary, we are required to minimize

$$\phi = \frac{\bar{a}^2}{T} \frac{(\bar{x}_4 - 2\bar{a}\bar{x}_3 + \bar{a}^2\bar{x}_2)}{\bar{x}_2\bar{x}_4 - \bar{x}_3^2} \tag{7.51}$$

with respect to

$$m \triangleq \frac{m_o}{m_1} \quad \text{and} \quad t \triangleq \frac{t_o}{T}$$

where

$$\bar{x}_2 \triangleq \bar{t}_1 + c^2\,\bar{t}_2 + m^2 t$$

$$\bar{x}_3 \triangleq \bar{t}_1 - c^3\,\bar{t}_2 + m^3 t$$

$$\bar{x}_4 \triangleq \bar{t}_1 + c^4\bar{t}_2 + m^4 t$$

$$\bar{t}_1 \triangleq \frac{t_1}{T} = \frac{(c - ct - mt)}{c + 1}$$

$$\bar{t}_2 \triangleq \frac{t_2}{T} = \frac{(1 - t + mt)}{c + 1}$$

$$c \triangleq \frac{m_2}{m_1}$$

$$\bar{a} \triangleq \frac{a}{m_1} \,.$$

As a numerical example, let

$$\bar{a} = \frac{a}{m_1} = 0.01$$

(i.e., the maximum acceleration from the rocket is 100 times greater than the designed acceleration for the accelerometer). Let us look at two particular cases:

Case 1:

$$c = \frac{m_2}{m_1} = 2$$

(i.e., the maximum braking thrust is twice the maximum rocket thrust).

Case 2:

$$c = 6 .$$

For the case, $c = 2$, the minimum value of ϕ is 0.75503 and occurs along the three sides of the allowable region at $t = 0$, $m = +1$, and $m = -c$. This means that the optimal input is $u = m$, for $\frac{2}{3}T$, and $u = -2m$, for $\frac{1}{3}T$ with no intermediate value of acceleration. As the value of c is increased, a local minimum ridge forms in the region shown in Fig. 7.4. For the case $c = 6$, the minimum value of ϕ is 0.69127 at $m = -2.7$ and $t = 0.2$. The optimal input is then $u = m_1$ for $0.763T$, $u = -6m_1$ for $0.037T$, and $u = 2.7m_1$ for $0.2T$.

The foregoing Example has two interesting features: (1) the optimal input may be designed without knowing the values of the parameters that are to be identified, and, (2) there is the possibility of one and only one intermediate thrust level. However, if the accelerometer is modelled by higher order terms in (7.28), then more intermediate values of thrust could be optimal.

D. EXAMPLE 2: A STABLE FIRST ORDER SYSTEM*

Find the optimal input to identify the parameter a in the first order system,

$$\dot{x} = -ax + au' , \qquad x(0) = 0 . \tag{7.53}$$

$$z = x + v$$

where

$$Ev(t)v(t') = r\delta(t - t') ,$$

* The first part of Example 2 was given in Nahi and Wallis [NA-1] .

and, the input is amplitude constrained by

$$|u'| \leq m . \tag{7.54}$$

The sensitivity equation is

$$\dot{\left(\frac{\partial x}{\partial a}\right)} = -a\left(\frac{\partial x}{\partial a}\right) - x + u' , \qquad \frac{\partial x}{\partial a}(0) = 0 . \tag{7.55}$$

By amplitude and time scaling, the system, sensitivity, and constraint equations become

$$\begin{aligned} \dot{x}_1 &= -x_1 + u, & x_1(0) &= 0 \\ \dot{x}_2 &= -x_1 - x_2 + u, & x_2(0) &= 0 \\ |u| &\leq 1 \end{aligned} \tag{7.56}$$

where the dot now denotes differentiation with respect to τ, and

$$\begin{aligned} \tau &\triangleq at \\ x_1 &\triangleq \frac{x}{m} \\ x_2 &\triangleq \frac{a}{m}\left(\frac{\partial x}{\partial a}\right) \\ u &\triangleq \frac{u'}{m} . \end{aligned} \tag{7.57}$$

The information "matrix" is simply the scalar

$$I = \int_0^T \frac{1}{r}\left(\frac{\partial x}{\partial a}\right)^2 dt = \frac{m^2}{a^3 r} \int_0^{T'} x_2^2 \, d\tau \tag{7.58}$$

and the variance is approximated by

$$P \cong I^{-1} = \frac{a^3 r}{m^2} x_3^{-1}(T) \tag{7.59}$$

ß

where

$$\dot{x}_3 = x_2^2, \qquad x_3(0) = 0 . \tag{7.60}$$

The input performance index is

$$\phi = P . \tag{7.61}$$

The gradient of ϕ with respect to the <u>k</u>th switch time t_k, is

$$\frac{\partial\phi}{\partial t_k} = -\frac{a^3 r}{m^2} x_3^{-2}(T)\frac{\partial x_3}{\partial t_k}(T) . \tag{7.62}$$

$$\frac{\partial x_3}{\partial t_k}(T)$$

is found by integrating (7.56) and

$$\begin{aligned}
\left(\frac{\partial \dot{x}_1}{\partial t_k}\right) &= -\left(\frac{\partial x_1}{\partial t_k}\right), & \frac{\partial x_1}{\partial t_k}(t_k) &= 2u(t_k^-) ; \\
\left(\frac{\partial \dot{x}_2}{\partial t_k}\right) &= -\left(\frac{\partial x_1}{\partial t_k}\right) - \left(\frac{\partial x_2}{\partial t_k}\right), & \frac{\partial x_2}{\partial t_k}(t_k) &= 2u(t_k^-) ; \\
\left(\frac{\partial \dot{x}_3}{\partial t_k}\right) &= 2x_2\left(\frac{\partial x_2}{\partial t_k}\right), & \frac{\partial x_3}{\partial t_k}(t_k) &= 0
\end{aligned} \tag{7.63}$$

from t_k to T.

Plots of the local minimum of the performance index for $N = 0$ to 3 are shown in Fig. 7.5. Figure 7.6 shows a plot of the global minimum of the performance index superimposed on a graph of the optimal switch times. As the length of the test increases, the center switch intervals become approximately equal.

Since the optimal input is piecewise constant (alternatively plus and minus one), (7.56) and (7.60) can readily be integrated from t_k to t_{k+1} to

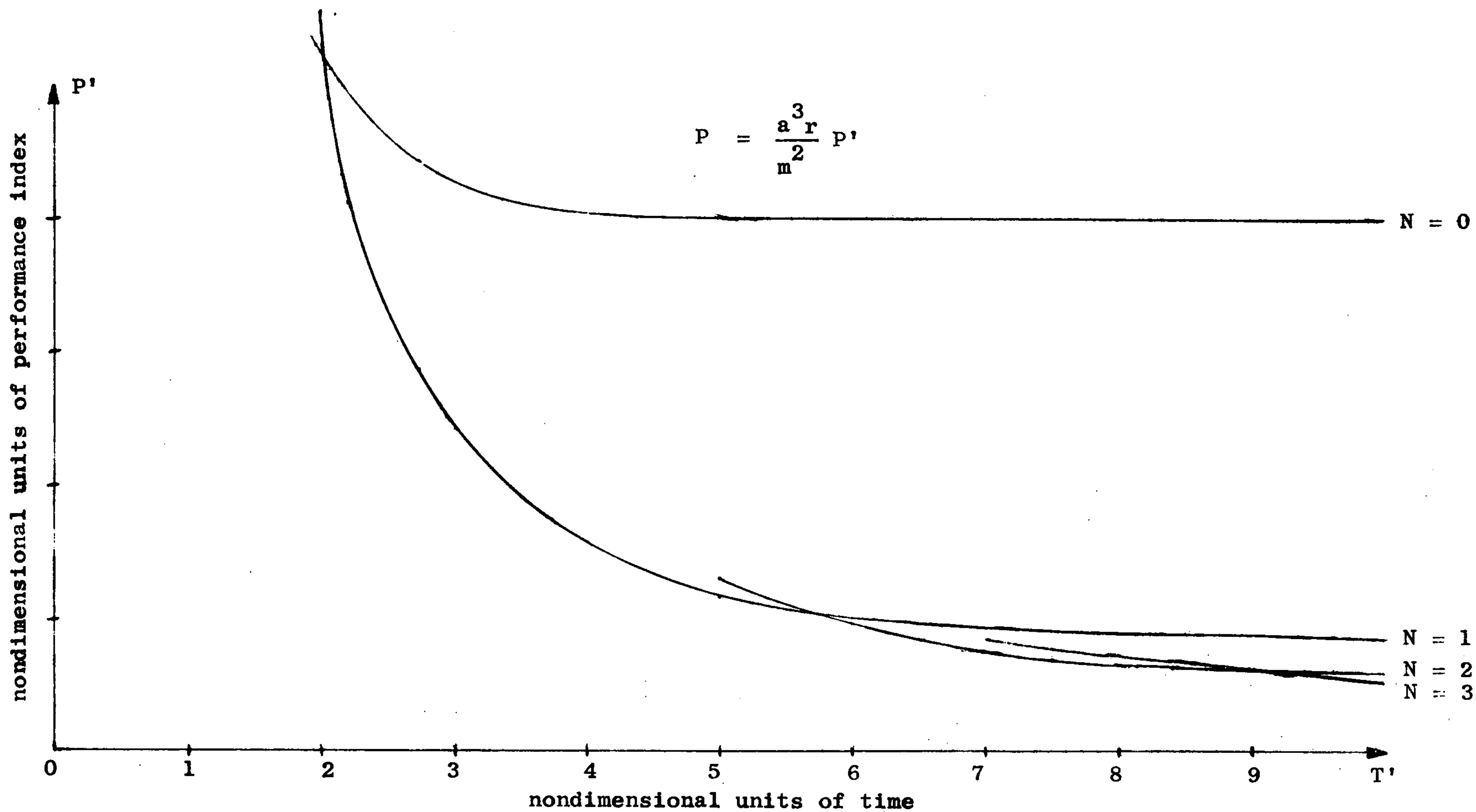


FIG. 7.5 PERFORMANCE INDEX VS LENGTH OF TEST

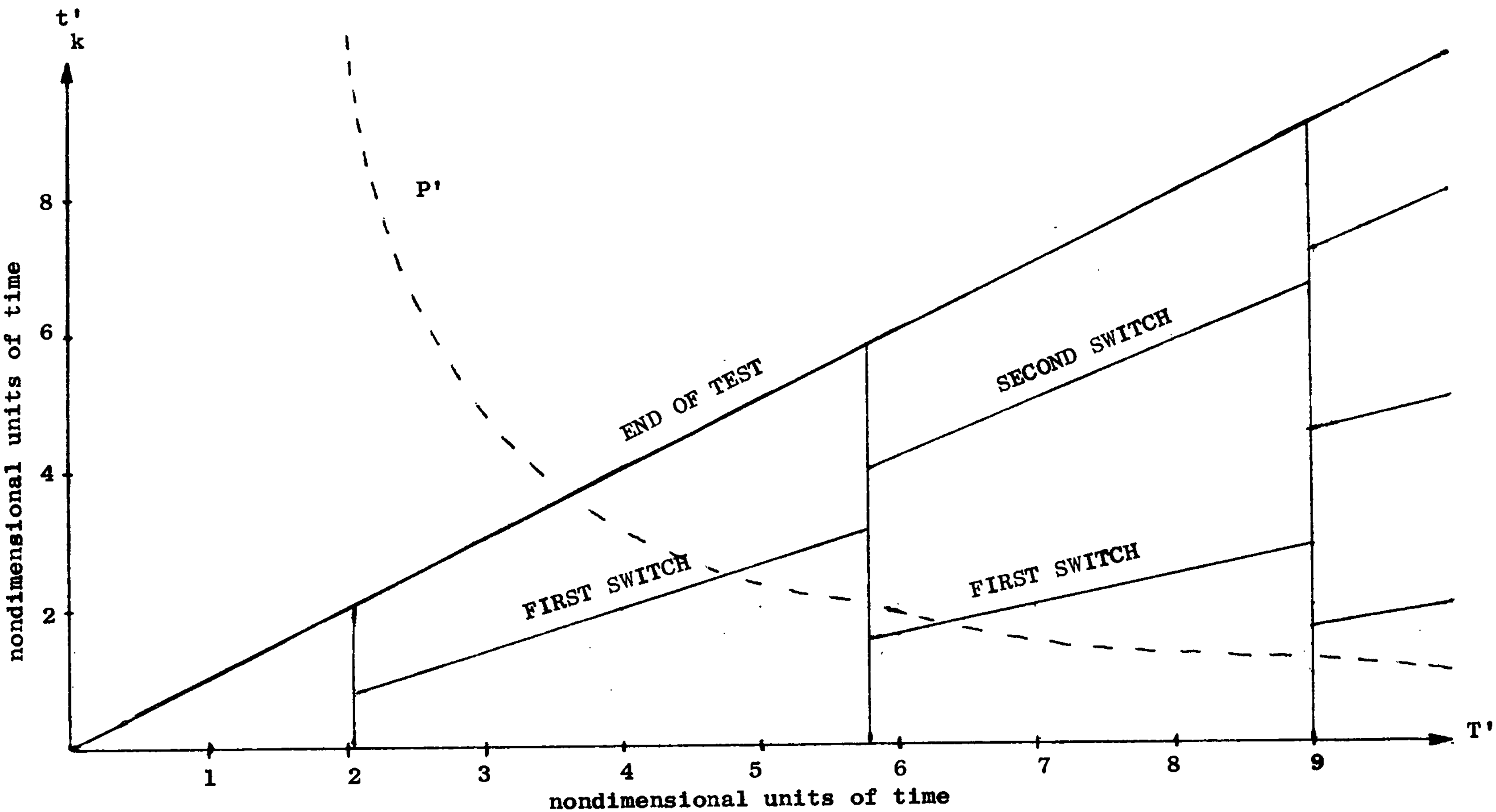


FIG. 7.6 SWITCH TIMES VS LENGTH OF TEST

give

$$
\begin{aligned}
x_1(k+1) &= u - [u - x_1(k)]e^{-\Delta k} \\
x_2(k+1) &= x_2(k)e^{-\Delta k} + [u - x_1(k)]\,\Delta_k e^{-\Delta k} \\
x_3(k+1) &= x_3(k) + \frac{1}{2}x_2^2(k)(1 - e^{-2\Delta k}) + \\
&\quad + \frac{1}{2}x_2(k)[u-x_1(k)][1 - e^{-2\Delta k}(2\Delta_k + 1)] \\
&\quad + \frac{1}{4}[u - x_1(k)]^2[1 - e^{-2\Delta k}(2\Delta_k^2 + 2\Delta_k + 1)]
\end{aligned}
\tag{7.64}
$$

where

$$\Delta_k = t_{k+1} - t_k .$$

An exact (square wave) analysis assuming all the intervals are equal can be made by using (7.64); where, for steady state it can be assumed that $x_1(k+1) = -x_1(k)$, $x_2(k+1) = -x_2(k)$ and they are negative for $u = -1$:

$$
\begin{aligned}
-x_{10} &= -1 - (-1 - x_{10})e^{-\Delta} \\
-x_{20} &= x_{20}e^{-\Delta} + (-1 - x_{10})\Delta e^{-\Delta} .
\end{aligned}
\tag{7.65}
$$

Hence,

$$
\begin{aligned}
x_{10} &= \frac{1 - e^{-\Delta}}{1 + e^{-\Delta}} \\
x_{20} &= \frac{2\Delta e^{-\Delta}}{(1 + e^{-\Delta})^2}
\end{aligned}
\tag{7.66}
$$

$$
x_3(T') = \frac{T'}{\Delta}\left\{\frac{1}{2}x_{20}^2(1 - e^{-2\Delta}) + \frac{1}{2}x_{20}(-1-x_{10})\left[1 - e^{-2\Delta}(2\Delta+1)\right] + \frac{1}{4}(1 + x_{10})^2\left[1 - e^{-2\Delta}(2\Delta^2+2\Delta+1)\right]\right\}
$$

Substituting for x_{10} and x_{20} and simplifying, we have

$$x_3(T') = \frac{T'}{\Delta(1 + e^{-\Delta})^3}\left[1 + (1-2\Delta)e^{-\Delta} - (1+2\Delta)e^{-2\Delta} - e^{-3\Delta}\right] . \qquad (7.67)$$

This has a maximum of $0.213\ T'$ at $\Delta = 3.28$. The corresponding angular frequency is $\omega = 0.958$.

If we had approximated this square wave with its first Fourier component, we would have

$$u = \frac{4}{\pi} \sin \omega\tau . \qquad (7.68)$$

Since

$$\frac{x_2}{u} = \frac{s}{(s + 1)^2} , \qquad (7.69)$$

the steady state amplitude ratio M is given by

$$M = \frac{\omega}{1 + \omega^2} . \qquad (7.70)$$

Thus,

$$x_3(T') = \int_0^{T'} x_2^2 \, d\tau = \left(\frac{4}{\pi}\right)^2 \frac{\omega^2}{(1+\omega^2)^2} \frac{T'}{\Delta} \int_0^{\Delta=\frac{\pi}{\omega}} \sin^2 \omega\tau \, d\tau$$

$$= \frac{8\omega^2 T'}{\pi^2 (1 + \omega^2)^2} . \qquad (7.71)$$

This has a maximum of $0.203\ T'$ at $\omega = 1$.

If we take the first two Fourier components of a square wave, we have

$$u = \frac{4}{\pi} \sin \omega t + \frac{4}{3\pi} \sin 3\omega\tau \tag{7.72}$$

$$x_2 = \frac{4}{\pi} \frac{\omega}{1+\omega^2} \sin \omega\tau + \frac{4}{3\pi} \frac{3\omega}{1+9\omega^2} \sin (3\omega\tau + \theta) \tag{7.73}$$

$$x_2^2 = \left(\frac{4}{\pi}\right)^2 \frac{\omega^2}{(1+\omega^2)^2} \sin^2 \omega\tau + \left(\frac{4}{3\pi}\right)^2 \frac{9\omega^2}{(1+9\omega^2)^2} \sin^2(3\omega\tau + \theta)$$
$$+ \left(\frac{4}{\pi}\right)^2 \frac{1}{3} \frac{3\omega^2}{(1+\omega^2)(1+9\omega^2)} \sin \omega\tau \sin (3\omega\tau + \theta) \tag{7.74}$$

$$x_3(T') = \frac{8\omega^2 T'}{\pi^2(1+\omega^2)^2} + \frac{8\omega^2 T'}{\pi^2(1+9\omega^2)^2} \tag{7.75}$$

This has a maximum of 0.211 T' at $\omega = 0.97$. The first three Fourier components yield a maximum of 0.212 T' at $\omega = 0.96$. Taking the first, second, or third Fourier components yield a very good approximation to the exact steady state solution. The computation is much simpler. In this case, we had to optimize with respect to only one parameter, ω.

E. EXAMPLE 3: A STABLE FIRST ORDER SYSTEM WITH PROCESS NOISE

Find the optimal input to identify a and K in the first order system:

$$\dot{x} = -ax + au' + w, \qquad x(0) = 0$$
$$z = x + v \tag{7.76}$$

where

$$Ew(t)\, w(t') = q\delta(t - t')$$

$$Ev(t)\, v(t') = r\delta(t - t')$$

and the input is amplituded constrained by $|u'| \leq m$. The steady state Kalman filter representation for this system is

$$\dot{\hat{x}} = -a\hat{x} + au' + K\nu, \qquad \hat{x}(0) = 0$$

$$\nu = z - \hat{x} \tag{7.77}$$

where

$$E\nu(t)\,\nu(t') = r\delta(t - t') .$$

The steady state covariance is given by

$$P = -ar + \sqrt{a^2 r^2 + qr} \tag{7.78}$$

hence, K is given by

$$K = -a + \sqrt{a^2 + q/r} . \tag{7.79}$$

Thus, we may identify the intensity of the process noise q by identifying the steady state gain K. For no process noise $q = 0 \Rightarrow K = 0$ which was considered in the previous Example.

The identification criterion is to minimize

$$J = \frac{1}{2}\int_0^T \frac{1}{r}(z - \hat{x})^2 dt \tag{7.80}$$

with respect to the unknown parameters a and K, subject to the constraint

$$\dot{\hat{x}} = -(a + K)\hat{x} + au' + Kz, \qquad \hat{x}(0) = 0 . \tag{7.81}$$

The first order sensitivity equations are

$$\left(\dot{\frac{\partial \hat{x}}{\partial a}}\right) = -(a + K)\left(\frac{\partial \hat{x}}{\partial a}\right) - \hat{x} + u', \qquad \frac{\partial \hat{x}}{\partial a}(0) = 0$$

$$\left(\dot{\frac{\partial \hat{x}}{\partial K}}\right) = -(a + K)\left(\frac{\partial \hat{x}}{\partial K}\right) - \hat{x} + z, \qquad \frac{\partial \hat{x}}{\partial K}(0) = 0 . \tag{7.82}$$

The information matrix for a and K is

$$I = \frac{1}{r}\begin{bmatrix} \int_o^T E\, x_2^2\, dt & \int_o^T E\, x_2 x_3\, dt \\ \int_o^T E\, x_2 x_3\, dt & \int_o^T E\, x_3^2\, dt \end{bmatrix} \tag{7.83}$$

where

$$\begin{aligned} \dot{x}_1 &= -ax_1 + au + K\nu, & x_1(0) &= 0 \\ \dot{x}_2 &= -(a+K)x_2 - x_1 + u', & x_2(0) &= 0 \\ \dot{x}_3 &= -(a+K)x_3 + \nu, & x_3(0) &= 0\,. \end{aligned} \tag{7.84}$$

Let $x = \bar{x} + \delta x$ so that

$$\begin{bmatrix} \dot{\bar{x}}_1 \\ \dot{\bar{x}}_2 \\ \dot{\bar{x}}_3 \end{bmatrix} = \begin{bmatrix} -a & 0 & 0 \\ -1 & -(a+K) & 0 \\ 0 & 0 & -(a+K) \end{bmatrix}\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} + \begin{bmatrix} a \\ 1 \\ 0 \end{bmatrix} u', \quad \begin{aligned} \bar{x}_1(0) &= 0 \\ \bar{x}_2(0) &= 0 \\ \bar{x}_3(0) &= 0 \end{aligned} \tag{7.85}$$

and

$$\begin{bmatrix} \delta\dot{x}_1 \\ \delta\dot{x}_2 \\ \delta\dot{x}_3 \end{bmatrix} = \begin{bmatrix} -a & 0 & 0 \\ -1 & -(a+K) & 0 \\ 0 & 0 & -(a+K) \end{bmatrix}\begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \delta x_3 \end{bmatrix} + \begin{bmatrix} K \\ 0 \\ 1 \end{bmatrix} \nu\,, \tag{7.86}$$

$$\delta x_1(0) = 0$$
$$\delta x_2(0) = 0 \qquad (7.86) \text{ Contd.}$$
$$\delta x_3(0) = 0 \; .$$

The expectations in the information matrix are given by

$$E\, x_2^2 = \bar{x}_2^2 + X_{22}$$
$$E\, x_2 x_3 = \bar{x}_2 \bar{x}_3 + X_{23} \qquad (7.87)$$
$$E\, x_3^2 = \bar{x}_3^2 + X_{33}$$

but $x_3 \equiv 0$, and X is given by solving

$$\dot{X}_{11} = -2a\, X_{11} + K^2 r, \qquad X_{11}(0) = 0$$
$$\dot{X}_{12} = -X_{11} - (2a + K)X_{12}, \qquad X_{12}(0) = 0$$
$$\dot{X}_{13} = -(2a + K)X_{13} + Kr, \qquad X_{13}(0) = 0$$
$$\dot{X}_{22} = -2X_{12} - 2(a + K)X_{22}, \qquad X_{22}(0) = 0 \qquad (7.88)$$
$$\dot{X}_{23} = -X_{13} - 2(a + K)X_{23}, \qquad X_{23}(0) = 0$$
$$\dot{X}_{33} = -2(a + K)X_{33} + r, \qquad X_{33}(0) = 0 \; .$$

The information matrix now becomes

$$I = \frac{1}{r} \begin{bmatrix} \int_o^T \bar{x}_2^2 + X_{22}\, dt & \int_o^T X_{23}\, dt \\ \int_o^T X_{23}\, dt & \int_o^T X_{33}\, dt \end{bmatrix} \qquad (7.89)$$

where for long tests, the covariance elements are approximately constant, so that

$$\int_o^T X_{22}dt \cong \frac{K^2 rT}{2a(2a + K)(a + K)}$$

$$\int_o^T X_{23}dt \cong \frac{-KrT}{2(2a + K)(a + K)} \qquad (7.90)$$

$$\int_o^T X_{33}dt \cong \frac{rT}{2(a + K)} \quad .$$

The lower bound of the covariance matrix for a and K is

$$P = r \frac{\begin{bmatrix} \int_o^T X_{33}\, dt & -\int_o^T X_{23}\, dt \\ -\int_o^T X_{23}\, dt & \int_o^T X_{22}dt + \int_o^T \bar{x}_2^2\, dt \end{bmatrix}}{\left[\int_o^T X_{22}dt + \int_o^T \bar{x}_2^2 dt\right] \cdot \int_o^T X_{33}\, dt - \left[\int_o^T X_{23}\, dt\right]^2} \qquad (7.91)$$

The optimal input to minimize the variance of a and/or K is found by maximizing

$$\int_o^T \bar{x}_2^2\, dt$$

or minimizing

$$\phi_d = \frac{1}{\int_o^T \bar{x}_2^2\, dt} \qquad (7.92)$$

subject to the constraints

$$\dot{\bar{x}}_1 = -a\bar{x}_1 + au', \qquad \bar{x}_1(0) = 0$$

$$\dot{\bar{x}}_2 = -\bar{x}_1 - (a + K)\bar{x}_2 + u', \qquad \bar{x}_2(0) = 0 \tag{7.93}$$

$$|u'| \leq m .$$

This reduces to the case without process noise if K is set equal to zero.

The optimal input continues to be bang-bang, but the switch times are changed by the addition of process noise. Normalizing the constraint equations, we have

$$\dot{x}_1 = -x_1 + u \qquad x_1(0) = 0$$

$$\dot{x}_2 = -x_1 - \eta x_2 + u, \qquad x_2(0) = 0 \tag{7.94}$$

$$|u| \leq 1$$

where the dot now denotes differentiation with respect to τ, and

$$\tau \triangleq at$$

$$x_1 \triangleq \frac{\bar{x}_1}{m}$$

$$x_2 \triangleq \frac{a}{m}\bar{x}_2 \tag{7.95}$$

$$u \triangleq \frac{u'}{m}$$

$$\eta \triangleq \frac{a + K}{a} = \sqrt{1 + \frac{q}{a^2 r}} .$$

The problem may now be solved as in the previous Example for different values of η. The performance index ϕ, is shown vs the test length for various values of η in Figs. 7.7 through 7.9. As the process noise increases, the switch intervals become shorter and the effectiveness of the input is reduced. Figure 7.10 shows the performance index and switch times for $\eta = 2$.

$$\eta \triangleq \sqrt{1 + \frac{q}{a^2 r}} \qquad P = \frac{a^3 r}{m^2} P'$$

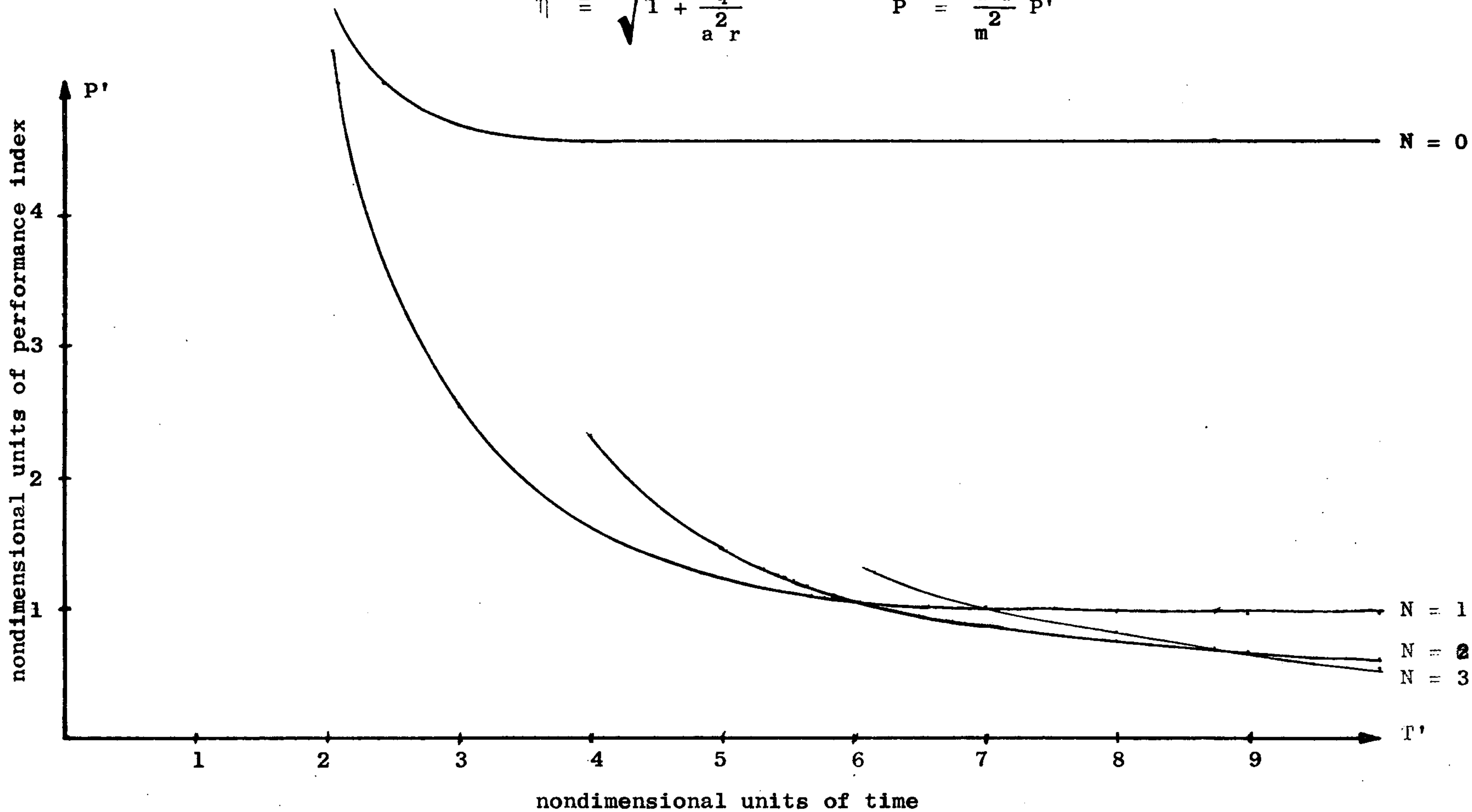


FIG. 7.7 PERFORMANCE INDEX VS LENGTH OF TEST FOR η = 1.1

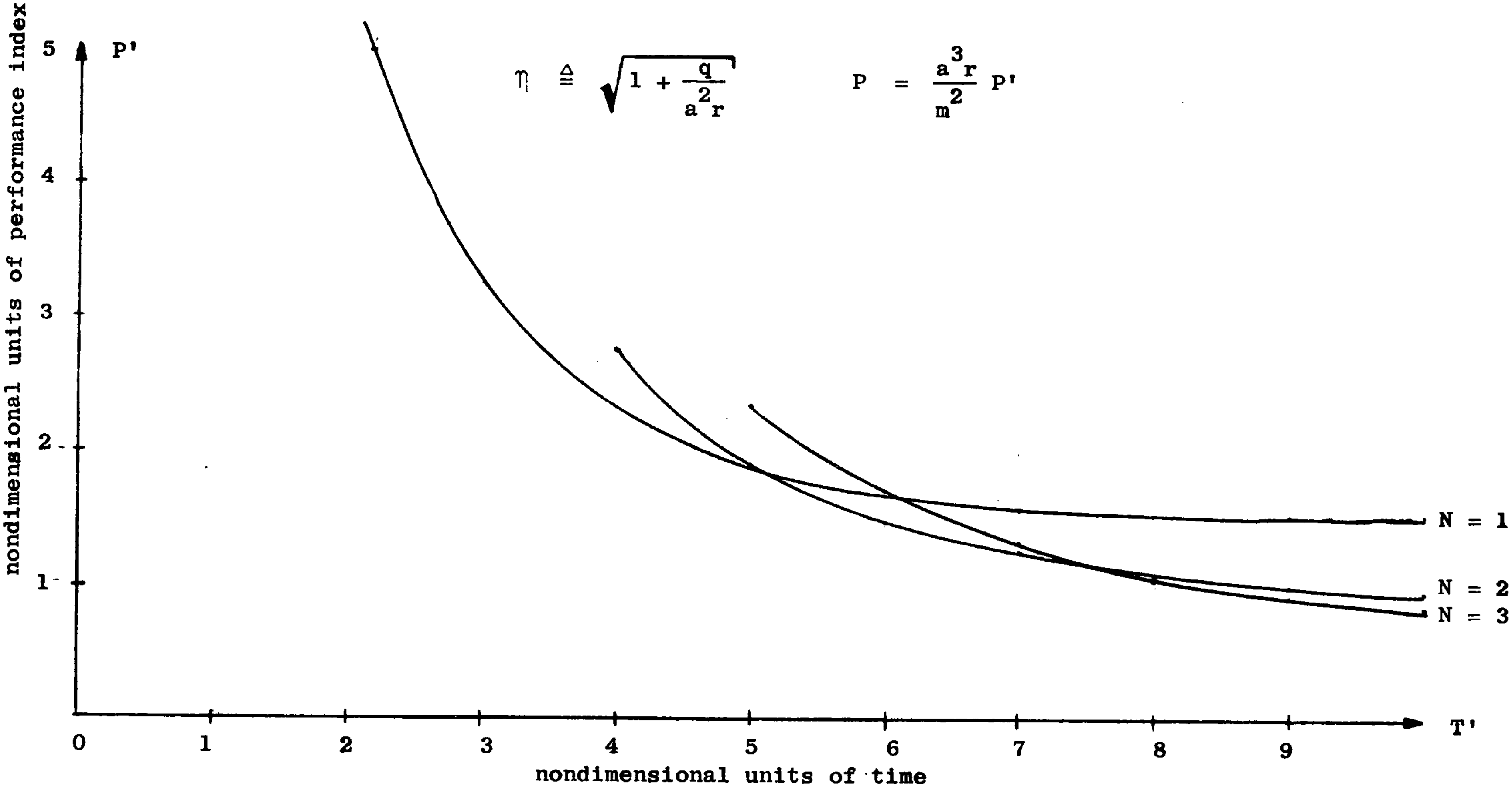


FIG. 7.8 PERFORMANCE INDEX VS LENGTH OF TEST FOR $\eta = 1.5$

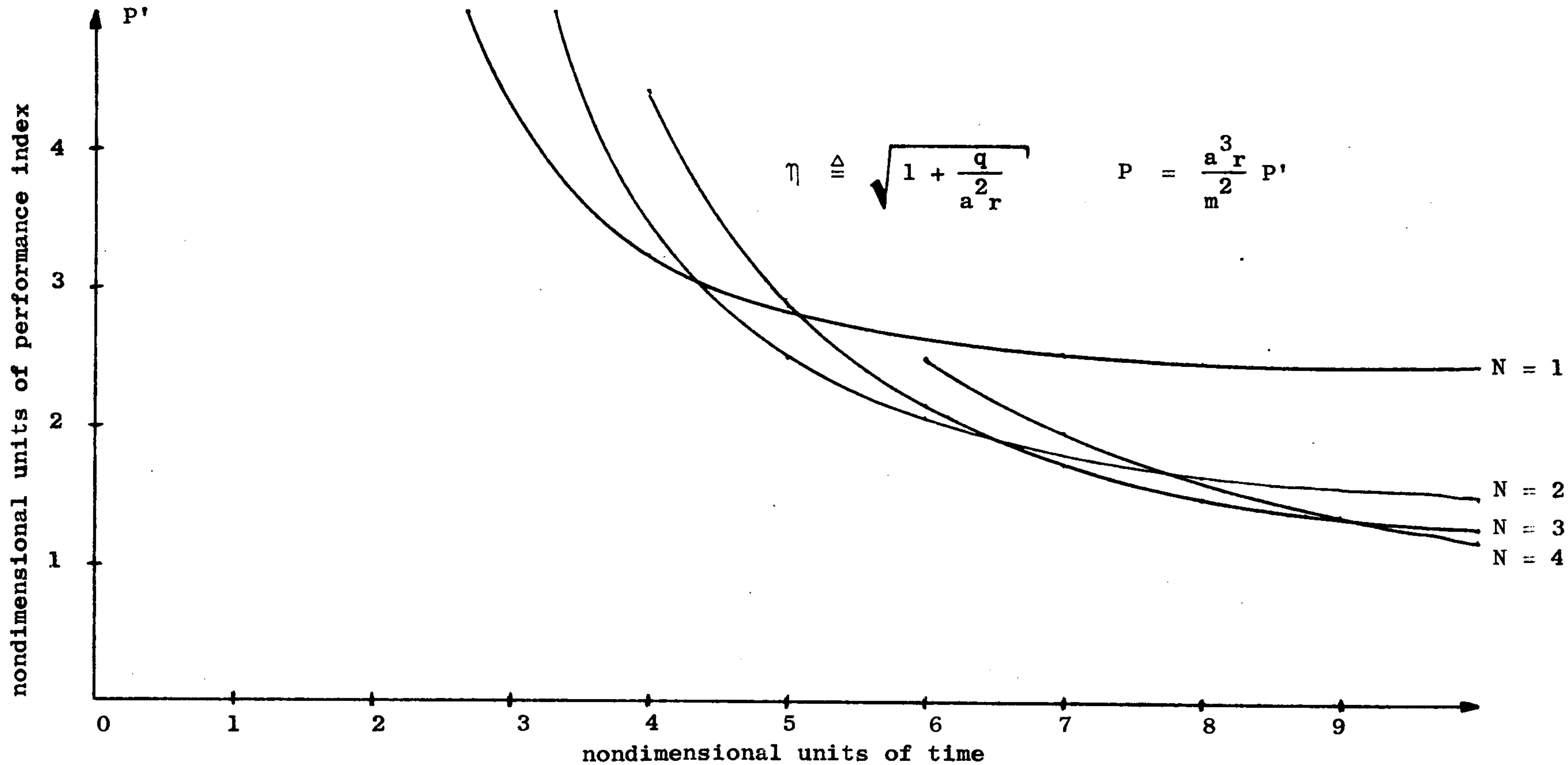


FIG. 7.9 PERFORMANCE INDEX VS LENGTH OF TEST FOR η = 2.0

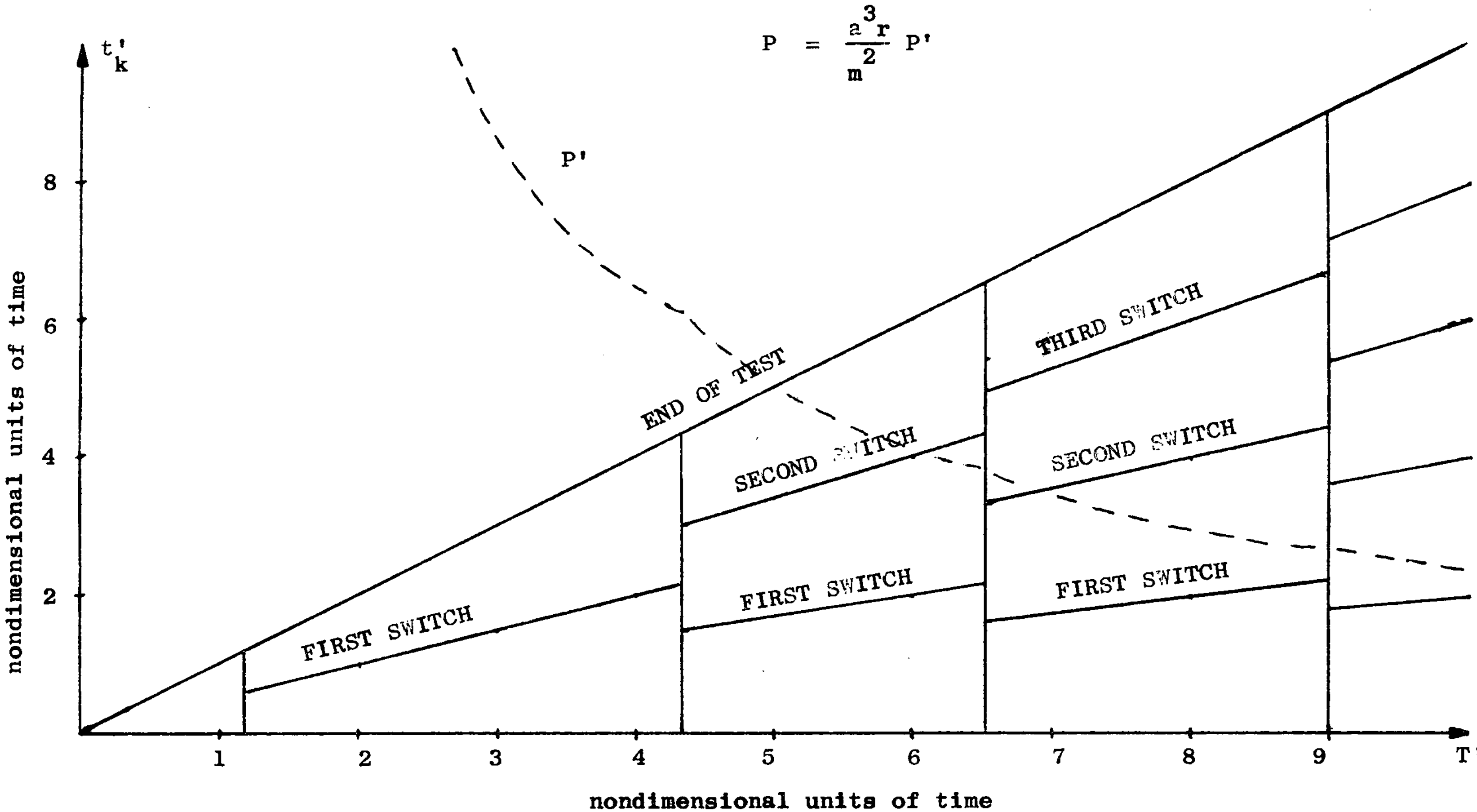


FIG. 7.10 SWITCH TIMES VS LENGTH OF TEST FOR $\eta = 2.0$

For long tests, we can show that the increased information from the covariance term is not sufficient to compensate for the lost effectiveness of the input (except where the deterministic input is severely constrained). For a long test, let us approximate the input with

$$u = \frac{4}{\pi} \sin \omega\tau \ . \tag{7.96}$$

Since

$$\frac{x_2}{u} = \frac{s}{(s+1)(s+\eta)} \tag{7.97}$$

the steady state amplitude ratio M is given by

$$M = \frac{\omega}{\sqrt{(1+\omega^2)(\eta^2+\omega^2)}} \tag{7.98}$$

and

$$\int_0^{T'} x_2^2 \, d\tau = \left(\frac{4}{\pi}\right)^2 \frac{\omega^2}{(1+\omega^2)(\eta^2+\omega^2)} \frac{1}{2} T' \ . \tag{7.99}$$

This has a maximum of

$$\frac{8T'}{\pi^2(\eta+1)^2} \quad \text{at} \quad \omega = \sqrt{\eta} \ .$$

If K is known, the covariance of a is

$$\phi = \frac{r \int_0^T X_{33} dt}{\left[\int_0^T X_{22} dt + \int_0^T \bar{x}_2^2 dt\right] \cdot \int_0^T X_{33} dt - \left[\int_0^T X_{23} dt\right]^2} \tag{7.100}$$

For long tests, the inverse of the covariance of a is

$$\frac{1}{\phi} = \frac{m^2}{a^3 r} \frac{8T'}{\pi^2(\eta + 1)^2} + \frac{(\eta - 1)^2 T'}{2a^2(\eta + 1)^2} . \qquad (7.101)$$

If we let

$$\alpha \triangleq \frac{q}{a^2 r} \quad \text{and} \quad \beta \triangleq \frac{m}{\sqrt{ar}} ,$$

then,

$$\frac{1}{\phi} = \frac{T'}{2a^2} \frac{\left[\frac{16}{\pi^2} \beta^2 + \left(\sqrt{\alpha + 1} - 1\right)^2\right]}{\left(\sqrt{\alpha + 1} + 1\right)^2} \qquad (7.102)$$

A plot of this function is shown in Fig. 7.11 for different values of β. From this figure, we can see that a little process noise usually degrades the overall accuracy of identification. However, where the input u is restricted to small values (β small), a larger amount of process noise can increase accuracy.

To get an idea of a reasonable amount of process noise compared with the deterministic input, let us assume that the process noise could be generated through the input u, $w = au$, and that we constrain the variance of u so that $3\sigma_u$ equals the magnitude of the inequality constraint:

$$\sigma_w = a\sigma_u \leq \frac{am}{3} . \qquad (7.103)$$

If the correlation time is μ, then

$$q = 2\mu \, \sigma_w^2 \leq \frac{2\mu a^2 m^2}{9} . \qquad (7.104)$$

In terms of α and β this inequality becomes

$$\alpha = \frac{q}{a^2 r} \leq \frac{2\mu a^2 m^2}{9a^2 r} = \frac{2\mu a}{9} \beta^2 \qquad (7.105)$$

so that a realistic α in Fig. 7.11 is very small and would only degrade the overall identification accuracy.

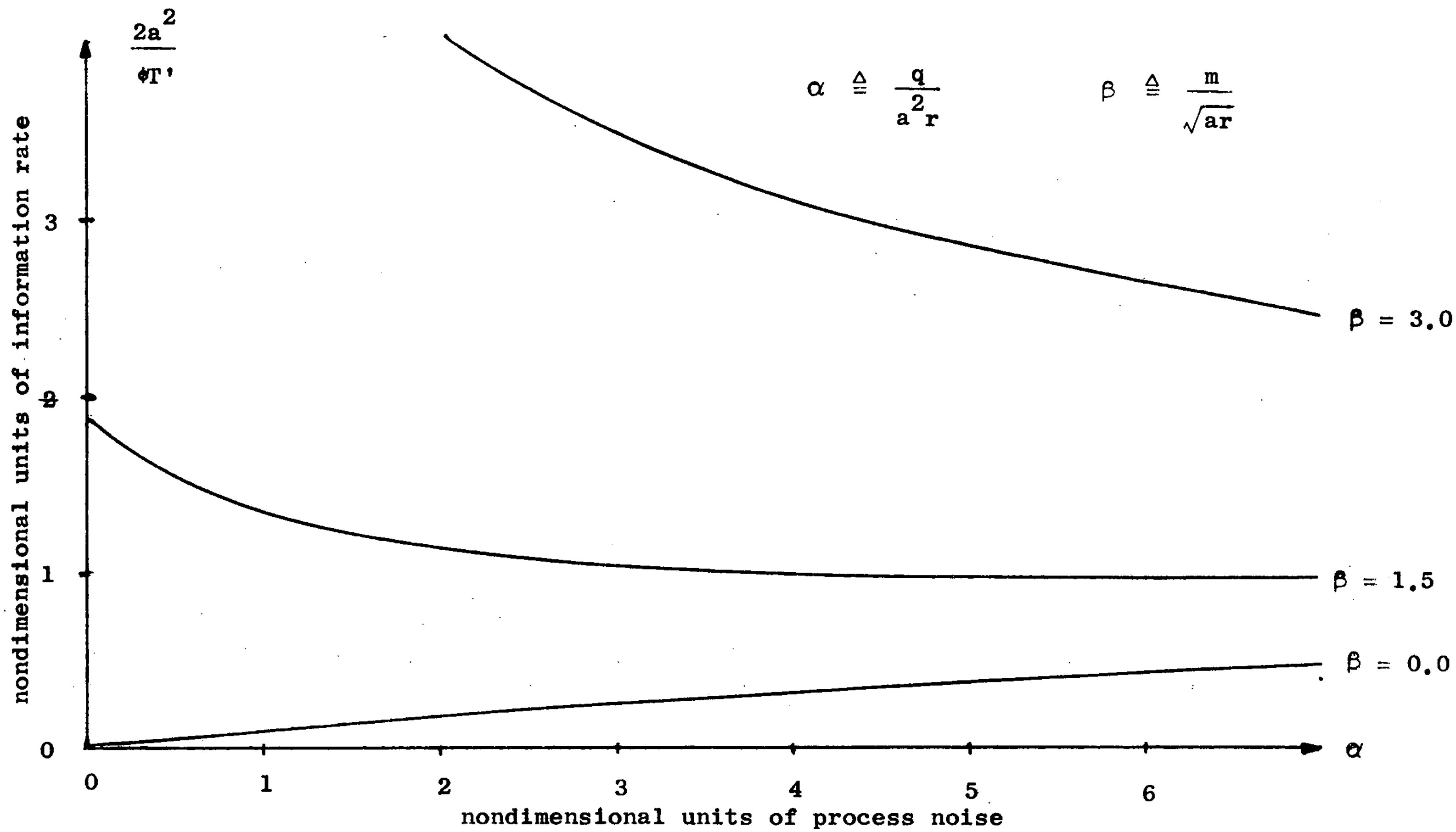


FIG. 7.11 RECIPROCAL OF PERFORMANCE INDEX VS AMOUNT OF PROCESS NOISE

F. EXAMPLE 4: A STABLE FIRST ORDER SYSTEM WITH A STATE INEQUALITY CONSTRAINT

Solve the problem in section 7.D with the addition of a state constraint

$$|x| \leqq \alpha m . \qquad (7.106)$$

For $\alpha \geqq 1$ this constraint has no effect since x is always within the region $|x| \leqq m$, for $|u| \leqq m$. The optimal solution is made up of control constrained arcs ($u' = \pm m$) and state constrained arcs ($x = \pm \alpha m$).

The scaled equations are

$$\begin{aligned}
\dot{x}_1 &= -x_1 + u & x_1(0) &= 0 \\
\dot{x}_2 &= -x_2 - x_1 + u, & x_2(0) &= 0 \\
\dot{x}_3 &= x_2^2 , & x_3(0) &= 0 \\
|u| &\leqq 1, & & \\
|x_1| &\leqq \alpha, & 0 < \alpha &< 1 .
\end{aligned} \qquad (7.107)$$

The time, t_a, needed to get to the state constrained arc is given by

$$\begin{aligned}
t_a &= -\ln \frac{1 - \alpha}{1 - x_1} \quad \text{if} \quad u = +1 \\
t_a &= -\ln \frac{1 - \alpha}{1 - x_1} \quad \text{if} \quad u = -1 .
\end{aligned} \qquad (7.108)$$

Let us define the switch times as the time when the control, u goes to ± 1. If the interval between switch times is greater than t_a, then we follow a constrained arc for a portion of the time between switch times. A typical input and output sequence is shown in Fig. 7.12. The problem may be solved as before with the addition that if the kth switch time is greater than $t_{k-1} + t_a$, then the control u is set equal to ± 1 from t_{k-1} to $t_{k-1} + t_a$ and set equal to $\pm \alpha$ from $t_a + t_{k-1}$ to t_k.

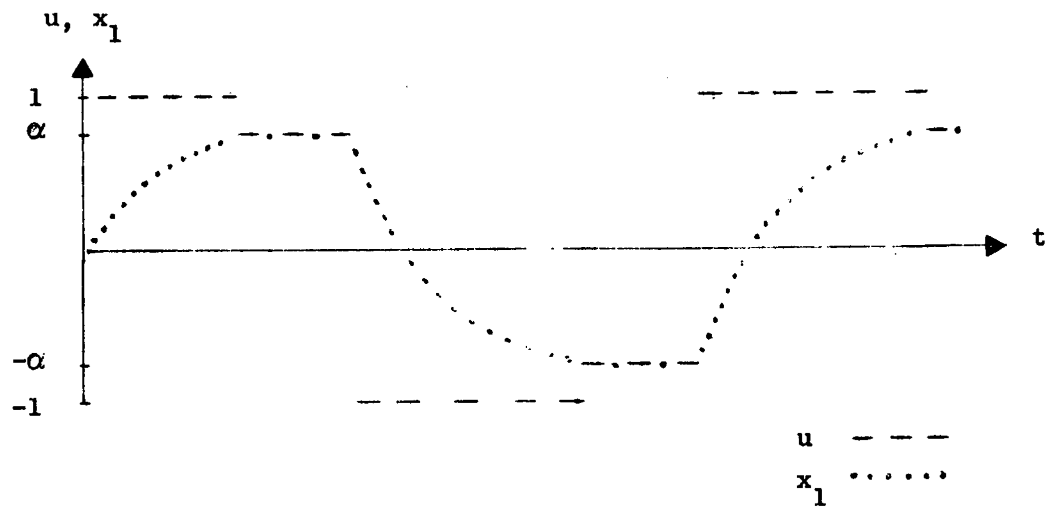


FIG. 7.12. INPUT AND OUTPUT CURVES WITH A STATE INEQUALITY CONSTRAINT.

This problem has been solved for $\alpha = \frac{1}{2}$ and is shown in Fig. 7.13. A comparison with Fig. 7.6 shows that the switch times are closer together than the case without a state constraint. As before, the first and last switch intervals are smaller than the central intervals and the central intervals are approximately equal.

Now let us see how the steady state solution is modified by the state inequality constraint. Recall that without a state-inequality constraint, the steady state solution yielded a time between switches of 3.28 time units, and that the maximum deviation in x was given by

$$x_{10} = \frac{1 - e^{-3.28}}{1 + e^{-3.28}} = 0.929 \qquad (7.109)$$

so that for $\alpha \geqq 0.929$, the steady solution is already solved with

$$\phi' = \frac{1}{0.213T'} = \frac{4.70}{T'} .$$

For $\alpha < 0.929$, we must allow for a portion of each switch interval to be on a state constrained arc.

Let us define t_c as the time between switches on a control constrained arc, and t_s the time on the state constraint. If we start

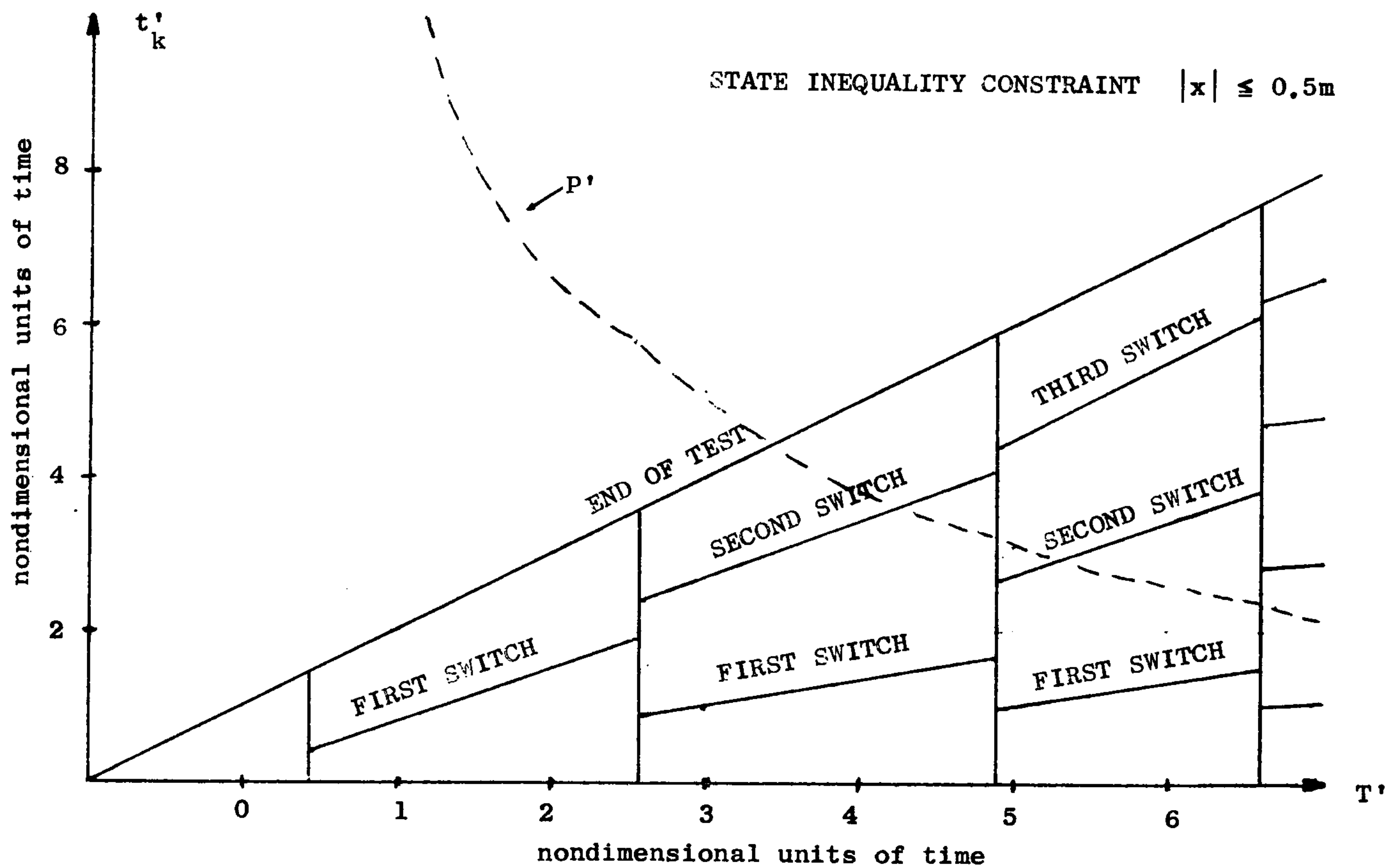


FIG. 7.13 SWITCH TIMES VS LENGTH OF TEST WITH A STATE INEQUALITY CONSTRAINT.

at $x_1(0) = -\alpha$ and $x_2(0) = -\beta$ with $u = +1$, then at t_c we have

$$x_1(t_c) = \alpha = (-\alpha - 1)e^{-t_c} + 1$$
$$x_2(t_c) = -\beta e^{-t_c} - (-\alpha - 1)t_c e^{t_c} . \qquad (7.110)$$

On the state constrained arc $x_1 = \alpha$ so that at $t_s + t_s$

$$x_1(t_c + t_s) = \alpha$$
$$x_2(t_c + t_s) = \beta = x_2(t_c)e^{-t_s} . \qquad (7.111)$$

The value of t_c and β is then

$$t_c = -\ln \frac{1 - \alpha}{1 + \alpha}$$
$$\beta = \frac{(1 - \alpha)t_c e^{-t_s}}{1 + \frac{1 - \alpha}{1 + \alpha} e^{-t_s}} . \qquad (7.112)$$

The total increase in x_3 during this time is $\Delta_c + \Delta_s$ where Δ_c is the increase of x_3 on the control constrained arc, and Δ_s is the increase on the state constrained arc. Δ_c and Δ_s are given by

$$\Delta_c = \frac{1}{2}\beta^2(1 - e^{-2t_c}) + \frac{1}{2} \beta(-\alpha-1)\left[1 - e^{-2t_c}(2t_c + 1)\right]$$
$$+ \frac{1}{4} (\alpha + 1)^2\left[1 - e^{-2t_c}(2t_c^2+2t_c+1)\right] \qquad (7.113)$$

$$\Delta_s = \frac{1}{2} x_2^2(t_c)(1 - e^{-2t_s}) .$$

The normalized covariance is then $\phi' = (t_c + t_s)/(\Delta_c + \Delta_s)T'$.

For a given value of α, this can be minimized with respect to t_s. A plot of t_c, $t_c + t_s$ and ϕ' T' is shown in Fig. 7.14 for $\alpha = 0.05$ to 1.0. As α becomes smaller, the covariance increases and the switch intervals become smaller.

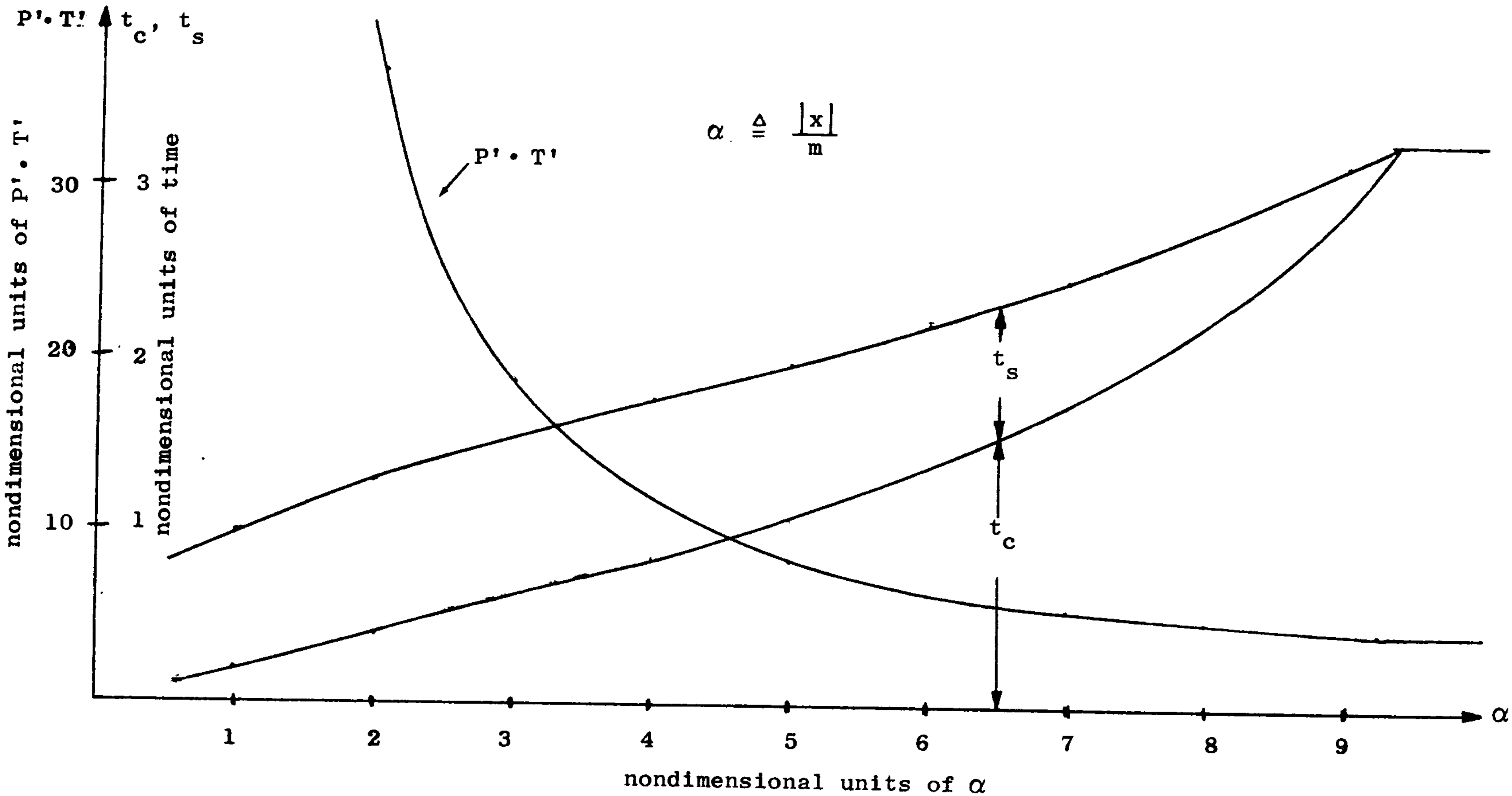


FIG. 7.14 PERFORMANCE INDEX VS MAGNITUDE OF STATE INEQUALITY CONSTRAINT.

The interesting feature of this Example is the fact that the state does stay on a state constraint for a portion of each cycle, and the frequency of switching is increased.

G. EXAMPLE 5: AN UNSTABLE FIRST ORDER SYSTEM

Find the optimal input to identify the parameter a in the unstable system

$$\begin{aligned} \dot{x} &= ax + au', \qquad x(0) = 0, \qquad a > 0 \\ z &= x + v . \end{aligned} \tag{7.114}$$

The sensitivity equation is

$$\dot{\left(\frac{\partial x}{\partial a}\right)} = a\left(\frac{\partial x}{\partial a}\right) + x + u . \tag{7.115}$$

If the only constraint is the input amplitude constraint $|u'| \leq m$, then the optimal input is $u' = \pm m$ with <u>no switching</u>. To maximize

$$\int_0^T \left(\frac{\partial x}{\partial a}\right)^2 dt,$$

we desire the largest possible x and u' terms driving the sensitivity equation. If $u' = \pm m$, then the input from u' and the input from x are as large (in absolute values) as possible. For this Example, whenever x is outside the region $|x| < m$, then the system cannot be controlled by an input whose amplitude is constrained by $|u'| \leq m$. For this reason we may wish to add a state variable inequality constraint

$$|x| \leq \alpha m \tag{7.116}$$

where $0 < \alpha < 1$.

The optimal input is made up of state constrained arcs $(u' = \mp \alpha m,$ $x = \pm \alpha m)$ and control constrained arcs $(u' = \pm m$ when $|x| \leqq \alpha m)$. By amplitude and time scaling, we have

$$
\begin{aligned}
\dot{x}_1 &= x_1 + u, & x_1(0) &= x_{10} \\
\dot{x}_2 &= x_2 + x_1 + u, & x_2(0) &= 0 \\
\dot{x}_3 &= x_2^2, & x_3(0) &= 0 \\
|u| &\leqq 1 \\
|x_1| &\leqq \alpha
\end{aligned}
\tag{7.117}
$$

where the dot now denotes differentiation with respect to τ, and

$$
\begin{aligned}
\tau &\triangleq at \\
x_1 &\triangleq \frac{x}{m} \\
x_2 &\triangleq \frac{a}{m}\left(\frac{\partial x}{\partial a}\right) \\
u &\triangleq \frac{u'}{m} .
\end{aligned}
$$

The information "matrix" is the scalar

$$
I = \int_0^T \left(\frac{\partial x}{\partial a}\right)^2 dt = \int_0^{T'} \left(\frac{m}{a}\right)^2 x_2^2 \frac{1}{a}\, d\tau = \frac{m^2}{a^3} x_3(T') \tag{7.118}
$$

so that the variance is

$$
P \cong \frac{a^3}{m^2} x_3^{-1} (T') . \tag{7.119}
$$

On a control constrained arc $(u = \pm 1)$, the solution to (7.117) is given by

$$
\begin{aligned}
x_1(\tau) &= (x_{10} + u)e^{\tau} - u \\
x_2(\tau) &= x_{20}e^{\tau} + (x_{10} + u)\tau e^{\tau} \\
x_3(\tau) &= x_{30} + \frac{1}{2} x_{20}^2(1 + e^{2\tau}) \\
&\quad + \frac{1}{2} x_{20}(x_{10} + u)[1 + e^{2\tau}(2\tau-1)] \\
&\quad + \frac{1}{4} (x_{10} + u)^2\left[e^{2\tau}(2\tau^2 - 2\tau + 1) - 1\right]
\end{aligned}
\tag{7.120}
$$

and on a state constrained arc $(x_1 = x_{10})$, the solution to (7.117) is given by

$$
\begin{aligned}
x_1(\tau) &= x_{10} \\
x_2(\tau) &= x_{20}e^{\tau} \\
x_3(\tau) &= x_{30} + \frac{1}{2} x_{20}^2(1 + e^{2\tau}) .
\end{aligned}
\tag{7.121}
$$

Let us now evaluate the performance index along two paths. The first path is $u = \mathrm{sgn}\ x(0)$ until the state constraint is hit and then to stay on the state constraint. The second path is $u = -\mathrm{sgn}\ x(0)$ until the other state constraint is hit and then to stay on that state constraint.

Along path 1 we have

$$
\begin{aligned}
x_1(\tau) &= (x_{10} + 1)e^{\tau} - 1 \\
x_2(\tau) &= (x_{10} + 1)\tau e^{\tau} \\
x_3(\tau) &= \frac{1}{4}(x_{10} + 1)^2\left[e^{2\tau}(2\tau^2 - 2\tau + 1) - 1\right],
\end{aligned}
\tag{7.122}
$$

for $0 \leqq \tau \leqq \tau_1$ and

$$
\begin{aligned}
x_1(\tau) &= \alpha \\
x_2(\tau) &= (x_{10} + 1)\tau_1 e^{\tau} \\
x_3(\tau) &= \frac{1}{4}(x_{10} + 1)^2 [e^{2\tau_1}(2\tau_1^2 - 2\tau_1 + 1) - 1] \\
&\quad + \frac{1}{2}(x_{10} + 1)^2 \tau_1^2 e^{2\tau_1}[1 + e^{2(\tau-\tau_1)}]
\end{aligned}
\tag{7.123}
$$

for $\tau \geqq \tau_1$, where τ_1 is given by

$$\tau_1 = \ell n \frac{1 + \alpha}{1 + x_{10}} . \tag{7.124}$$

Along path 2 we have

$$
\begin{aligned}
x_1(\tau) &= (x_{10} - 1)e^{\tau} + 1 \\
x_2(\tau) &= (x_{10} - 1)\tau e^{\tau} \\
x_3(\tau) &= \frac{1}{4}(x_{10} - 1)^2 [e^{2\tau}(2\tau^2 - 2\tau + 1) - 1]
\end{aligned}
\tag{7.125}
$$

for $0 \leqq \tau \leqq \tau_2$ and

$$
\begin{aligned}
x_1(\tau) &= -\alpha \\
x_2(\tau) &= (x_{10} - 1)\tau_2 e^{\tau} \\
x_3(\tau) &= \frac{1}{4}(x_{10} - 1)^2 [e^{2\tau_2}(2\tau_2^2 - 2\tau_2 + 1) - 1] \\
&\quad + \frac{1}{2}(x_{10} - 1)^2 \tau_2^2 e^{2\tau_2}(1 + e^{2(\tau-\tau_2)})
\end{aligned}
\tag{7.126}
$$

for $\tau \geqq \tau_2$, where τ_2 is given by

$$\tau_2 = \ell n \frac{1 + \alpha}{1 - x_{10}}$$

It can be verified that $x_3(\tau_1)$ along path 1 is greater than $x_3(\tau)$ along path 2. However, it can also be verified that $|x_2(\tau_2)|$ along path 2 is greater than $|x_2(\tau_2)|$ along path 1. This means that if $x_3(\tau_2)$ along path 2 is not greater than $x_3(\tau_2)$ along path 1, it will be at some later time. Let us designate τ_3 as the time at which $x_3(\tau_3)$ along both paths are equal. Also consider that once on a state constrained arc (with a sufficient magnitude for x_2), it is better to stay on that arc than go to the other constraint or go off and return to that arc.

For these reasons, we can say that the optimal input is $u = m \operatorname{sgn} x_o$ until the state constraint is hit and then is such as to stay on the constrained arc, for a test whose length is less than τ_3. However, for a test whose length is greater than τ_3, the optimal input is $u = -m \operatorname{sgn} x_o$ until the opposite constrained arc is hit and then is such as to stay on that constrained arc. However, in both cases, the optimal input involves going to a constrained arc and staying on the constraint.

This Example has two interesting features: (1) Since the state and sensitivity equations are unstable, the information matrix grows much faster than for a stable system. This means that an unstable system may be identified more accurately than a stable system. (2) The optimal input involves no switching.

H. EXAMPLE 6: AN UNSTABLE FIRST ORDER SYSTEM WITH TWO PARAMETERS

Find the optimal input to identify a and b of the first order system

$$\dot{x} = ax + bu', \qquad x(0) = 0, \qquad a > 0$$
$$z = x + v \tag{7.128}$$

with an input amplitude constraint $|u'| \leq m$. The two sensitivity equations are

$$\left(\frac{\partial \dot{x}}{\partial a}\right) = a\left(\frac{\partial x}{\partial a}\right) + x$$

$$\left(\frac{\partial \dot{x}}{\partial b}\right) = a\left(\frac{\partial x}{\partial b}\right) + u' \tag{7.129}$$

and the information matrix is

$$I = \frac{1}{r}\begin{bmatrix} \int_o^T \left(\frac{\partial x}{\partial a}\right)^2 dt & \int_o^T \left(\frac{\partial x}{\partial a}\right)\left(\frac{\partial x}{\partial b}\right) dt \\ \int_o^T \left(\frac{\partial x}{\partial a}\right)\left(\frac{\partial x}{\partial b}\right) dt & \int_o^T \left(\frac{\partial x}{\partial b}\right)^2 dt \end{bmatrix} \tag{7.130}$$

By amplitude and time scaling, we have

$$\begin{aligned} \dot{x}_1 &= x_1 + u, & x_1(0) &= 0 \\ \dot{x}_2 &= x_2 + x_1, & x_2(0) &= 0 \\ \dot{x}_3 &= x_3 + u, & x_3(0) &= 0 \\ |u| &\leq 1 \end{aligned} \tag{7.131}$$

where a dot now denotes differentiation with respect to τ, and

$$\begin{aligned} \tau &= at \\ x_1 &= \frac{a}{bm} x \\ x_2 &= \frac{a^2}{bm}\left(\frac{\partial x}{\partial a}\right) \\ x_3 &= \frac{a}{m}\left(\frac{\partial x}{\partial b}\right) . \end{aligned} \tag{7.132}$$

For the initial condition given, $x_3 \equiv x_1$. The information matrix becomes

$$I = \frac{m^2}{a^3 r} \begin{bmatrix} \frac{b^2}{a^2} x_5(T') & \frac{b}{a} x_6(T') \\ \frac{b}{a} x_6(T') & x_4(T') \end{bmatrix} \qquad (7.133)$$

where

$$\begin{aligned} \dot{x}_4 &= x_1^2 & x_4(0) &= 0 \\ \dot{x}_5 &= x_2^2 & x_5(0) &= 0 \\ \dot{x}_6 &= x_1 x_2, & x_6(0) &= 0 \,. \end{aligned} \qquad (7.134)$$

The covariance matrix is approximated by

$$P \cong \frac{ra^3}{m^2 b^2} \frac{\begin{bmatrix} a^2 x_4 & -abx_6 \\ -abx_6 & b^2 x_4 \end{bmatrix}}{x_4 x_5 - x_6^2} \,. \qquad (7.135)$$

If we weigh the coefficients of variation σ_a/a and σ_b/b equally, then our performance index becomes

$$\phi = \frac{x_4(T') + x_5(T')}{x_4(T')\, x_5(T') - x_6^2(T')} \,. \qquad (7.136)$$

Figures 7.15 through 7.17 show plots of the performance index versus one switch time for tests of $T' = 0.5$, 1.0, and 3.0 time units. In each case, one switch is better than no switches.

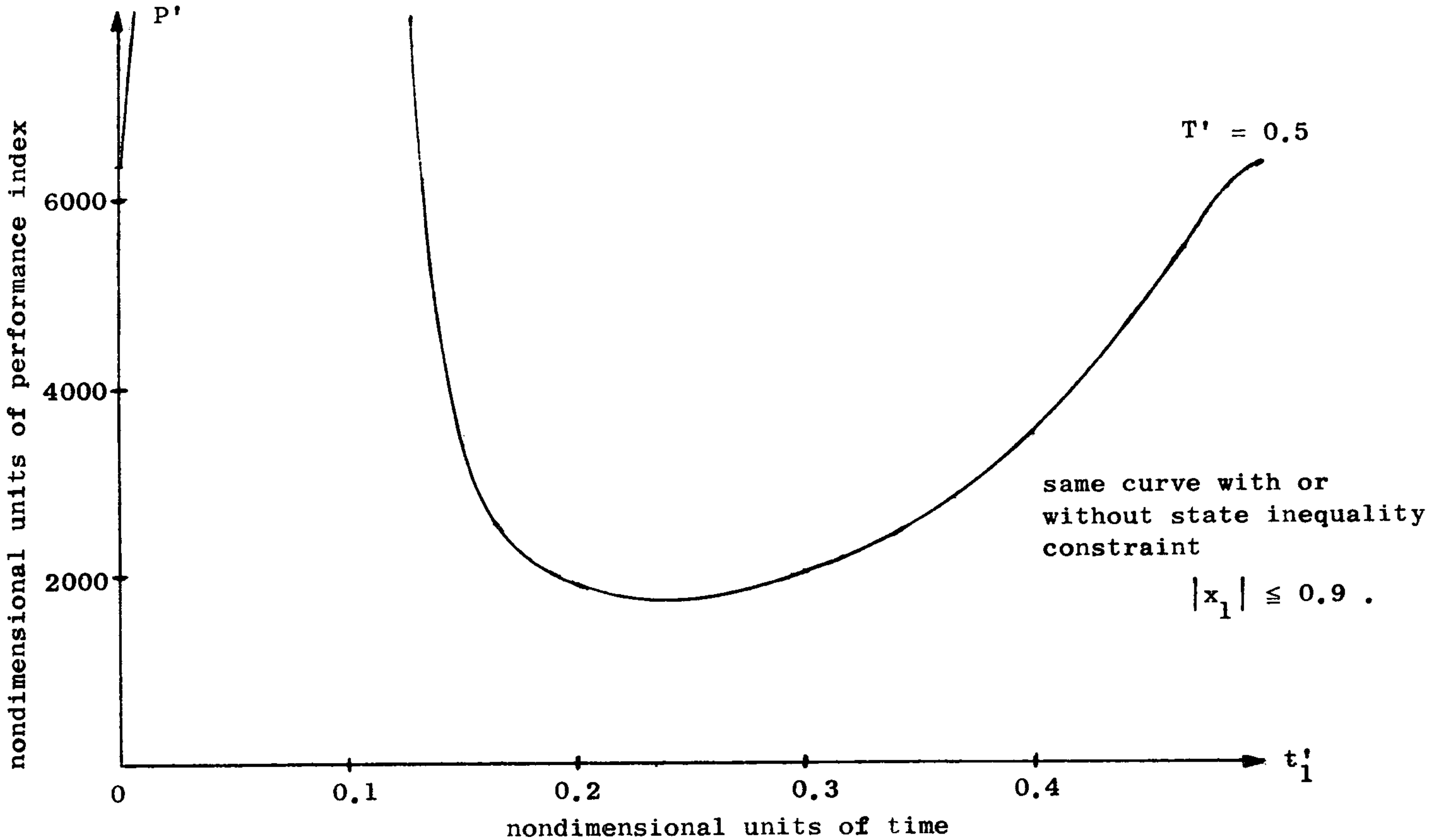


FIG. 7.15 PERFORMANCE INDEX VS ONE SWITCH TIME. T' = 0.5

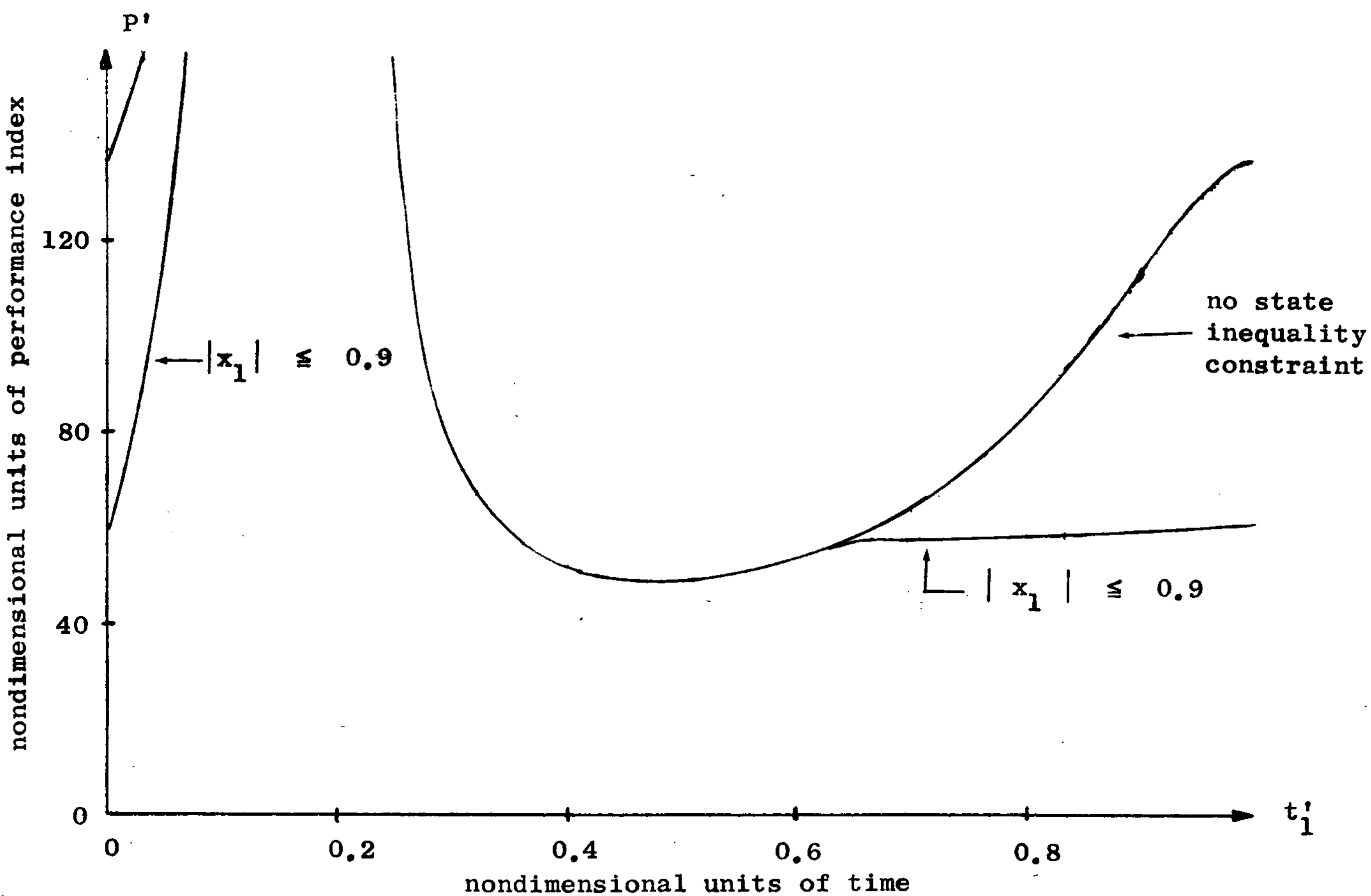


FIG. 7.16 PERFORMANCE INDEX VS ONE SWITCH TIME
$T' = 1.0$

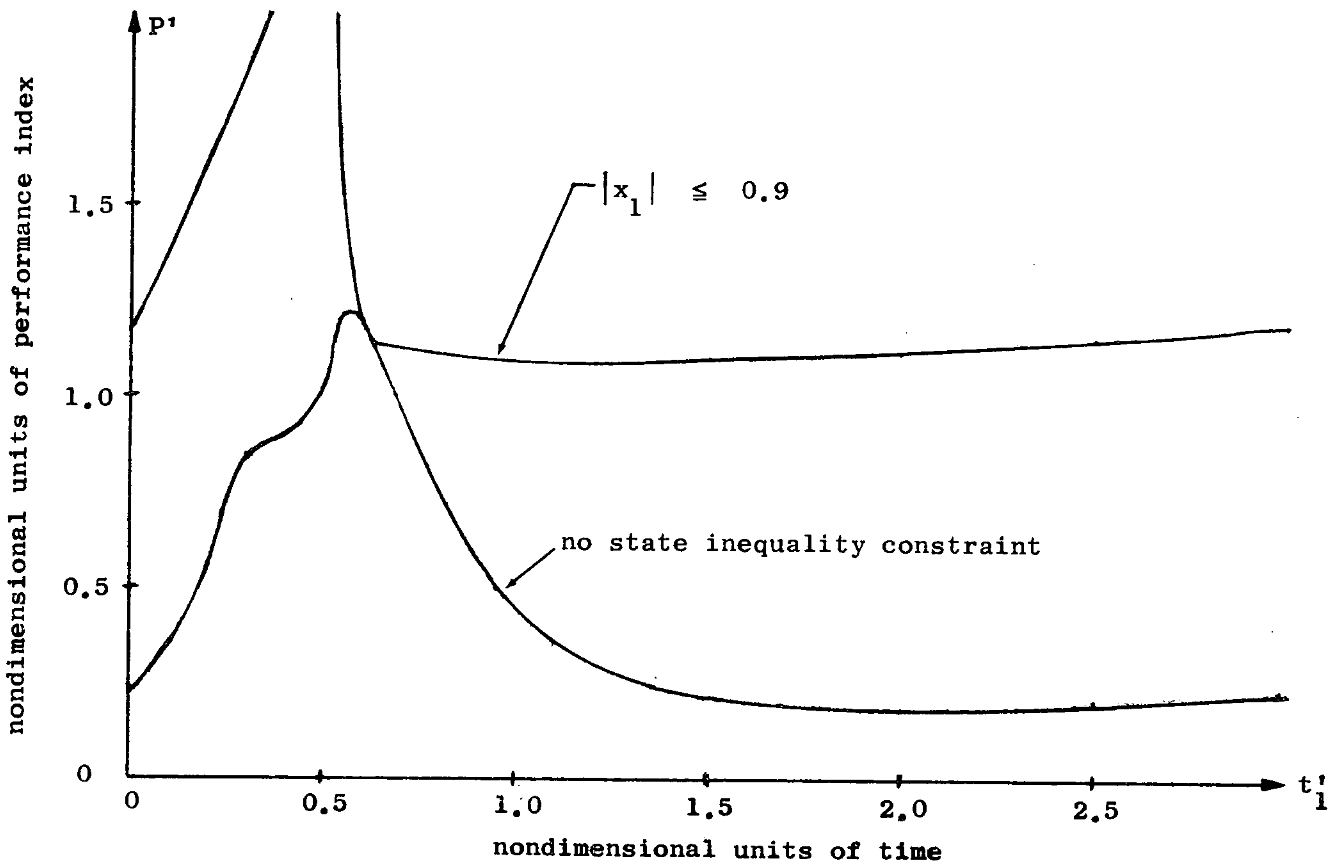


FIG. 7.17 PERFORMANCE INDEX VS ONE SWITCH TIME.
T' = 3.0

Plots of the performance index versus two switch times were also run. However, in each case the best two-switch sequence was the one-switch case. For this reason, it is believed that the optimal input is bang-bang with one and only one switch. This is in marked constrast with our stable systems that involve repeated switching for long tests.

Now let us solve the problem with the first order state-inequality constraint

$$|x| \leqq \alpha \frac{bm}{a}$$

where $0 < \alpha < 1$.

As in the previous Example (G), the optimal input is made up of state constrained

$$x = \pm \alpha \frac{bm}{a} ,$$

and control constrained

$$u' = \pm m$$

arcs.

In mechanizing a program to calculate the performance index as a function of the switch times, the switch times are defined as the times when the control u goes to +1 or to -1 (not when it goes to some intermediate value to stay on a state constraint).

Figures 7.15 through 7.17 also show plots of the performance index for the case $\alpha = 0.9$. As in the case without a state inequality constraint, one and only one switch is optimal. This example is quite similar to the previous unstable system. The main difference is that to identify two parameters, the optimal input involved one and only one switch.

Chapter VIII

OPTIMAL INPUT FOR THE IDENTIFICATION OF THE LONGITUDINAL DYNAMIC STABILITY DERIVATIVES

A. PROBLEM FORMULATION

The approximate longitudinal equations of motion (short-period oscillation) for an airplane are*

$$\dot{q} = \frac{M_{\dot{\alpha}} + M_q}{I_y} q + \left(\frac{M_{\dot{\alpha}} z_\alpha}{I_y m u_o} + \frac{M_\alpha}{I_y}\right)\alpha + \left(\frac{M_{\dot{\alpha}} z_{\delta_e}}{I_y m u_o} + \frac{M_{\delta_e}}{I_y}\right)\delta_e$$

$$\dot{\alpha} = q + \frac{z_\alpha}{m u_o}\alpha + \frac{z_{\delta_e}}{m u_o}\delta_e \tag{8.1}$$

where

q = pitch rate

α = angle of attack

δ_e = elevator deflection.

Let us assume that all the parameters except $M_{\dot{\alpha}}$ and M_q can be determined from wind tunnel tests. Hence, we wish to identify the normalized parameters

$$p_1 = \frac{M_{\dot{\alpha}}}{I_y} \text{ and } p_2 = \frac{M_q}{I_y}$$

from a flight test.

---

* This form of the equations was taken from Denery [DE-1]

For this test, let us assume that the only measurement is the pitch rate q, which is corrupted by white gaussian noise of density R (a scalar). Our problem is to determine the optimal input δ_e for the identification test with the constraint

$$|\delta_e| \leq \delta_{e_{max}} . \tag{8.2}$$

The identification performance index is

$$J = \frac{1}{2R} \int_o^T (z - q)^2 dt \tag{8.3}$$

so that the information matrix is

$$I_a = \frac{1}{R} \begin{bmatrix} \int_o^T \left(\frac{\partial q}{\partial p_1}\right)^2 dt & \int_o^T \left(\frac{\partial q}{\partial p_1}\right)\left(\frac{\partial q}{\partial p_2}\right) dt \\ \int_o^T \left(\frac{\partial q}{\partial p_1}\right)\left(\frac{\partial q}{\partial p_2}\right) dt & \int_o^T \left(\frac{\partial q}{\partial p_2}\right)^2 dt \end{bmatrix} . \tag{8.4}$$

If we approximate the covariance matrix for p_1 and p_2 by I_a^{-1}, and put an equal weighting on their accuracy, our input performance index becomes

$$\phi = \text{Tr } I_a^{-1} . \tag{8.5}$$

In order to evaluate the information matrix, we must calculate the two sets of sensitivity equations

$$\left(\frac{\dot{\partial q}}{\partial p_1}\right) = \frac{M_{\dot\alpha} + M_q}{I_y}\left(\frac{\partial q}{\partial p_1}\right) + \left(\frac{M_{\dot\alpha} z_\alpha}{I_y m u_o} + \frac{M_\alpha}{I_y}\right)\left(\frac{\partial \alpha}{\partial p_1}\right) + q + \frac{z_\alpha}{m u_o}\alpha$$
$$+ \frac{z_{\delta_e}}{m u_o}\delta_e, \qquad \frac{\partial q}{\partial p_1}(0) = 0 ; \tag{8.6}$$

$$\left(\frac{\partial \alpha}{\partial p_1}\right)^{\cdot} = \left(\frac{\partial q}{\partial p_1}\right) + \frac{z_\alpha}{mu_o}\left(\frac{\partial \alpha}{\partial p_1}\right), \qquad \frac{\partial \alpha}{\partial p_1}(0) = 0 \tag{8.6 cont.}$$

$$\left(\frac{\partial q}{\partial p_2}\right)^{\cdot} = \frac{M_{\dot\alpha} + M_q}{I_y}\left(\frac{\partial q}{\partial p_2}\right) + \left(\frac{M_{\dot\alpha} z_\alpha}{I_y mu_o} + \frac{M_\alpha}{I_y}\right)\left(\frac{\partial \alpha}{\partial p_2}\right) + q, \qquad \frac{\partial q}{\partial p_2}(0) = 0$$

$$\left(\frac{\partial \dot\alpha}{\partial p_2}\right) = \left(\frac{\partial q}{\partial p_2}\right) + \frac{z_\alpha}{mu_o}\left(\frac{\partial \alpha}{\partial p_2}\right), \qquad \frac{\partial \alpha}{\partial p_2}(0) = 0 \tag{8.7}$$

B. NORMALIZATION

The state, sensitivity, and constraint equations may also be written in the form:

$$\begin{aligned}
\dot q &= k_{32} q + k_{34}\alpha + \delta g_{31}\delta_e, & q(0) &= 0\\
\dot\alpha &= q + k_{54}\alpha, & \alpha(0) &= 0\\
\dot q^{(1)} &= k_{32} q^{(1)} + k_{34}\alpha^{(1)} + q + k_{54}\alpha, & q^{(1)}(0) &= 0\\
\dot\alpha^{(1)} &= q^{(1)} + k_{54}\alpha^{(1)}, & \alpha^{(1)}(0) &= 0\\
\dot q^{(2)} &= k_{32} q^{(2)} + k_{34}\alpha^{(2)} + q, & q^{(2)}(0) &= 0\\
\dot\alpha^{(2)} &= q^{(2)} + k_{54}\alpha^{(2)}, & \alpha^{(2)}(0) &= 0\\
|\delta_e| &\le \delta_{e\max},
\end{aligned} \tag{8.8}$$

where the ith superscript denotes the sensitivity equation for p_i. The correspondence between old and new coefficients is shown in Table 8.1. k_{32} and k_{34} are unknown and δg_{31} and k_{54} are known from wind tunnel testing. By amplitude and time scaling, we can reduce the above set of equations to the form:

Table 8.1

RELATIONSHIP BETWEEN COEFFICIENTS IN (8.1) and (8.8). The numerical values are those in [DE-1] for the C-8 airplane in a landing configuration. We assume that some parameters are known or unknown from wind tunnel testing.

Old Coefficients	New Coefficients	Numerical Values	Known or Unknown
$\frac{M_{\dot{\alpha}} + M_q}{I_y}$	k_{32}	-1.588	unknown
$\frac{M_{\dot{\alpha}} z_{\alpha}}{I_y m u_o} + \frac{M_{\alpha}}{I_y}$	k_{34}	-0.562	unknown
$\frac{M_{\dot{\alpha}} z_{\delta e}}{I_y m u_o} + \frac{M_{\delta e}}{I_y}$	δg_{31}	-1.658	known (if $\delta g_{51} = 0$)
$\frac{z_{\alpha}}{m u_o}$	k_{54}	-1.737	known
$\frac{z_{\delta e}}{m u_o}$	δg_{51}	0.005	assume = 0

$$
\begin{aligned}
\dot{x}_1 &= -c_1x_1 - c_2x_2 + u\,, & x_1(0) &= 0 \\
\dot{x}_2 &= x_1 - x_2\,, & x_2(0) &= 0 \\
\dot{x}_3 &= -c_1x_3 - c_2x_4 + x_1 - x_2\,, & x_3(0) &= 0 \\
\dot{x}_4 &= x_3 - x_4\,, & x_4(0) &= 0 \\
\dot{x}_5 &= -c_1x_5 - c_2x_6 + x_1\,, & x_5(0) &= 0 \\
\dot{x}_6 &= x_5 - x_6\,, & x_6(0) &= 0 \\
|u| &\leq 1\,, & &
\end{aligned}
\qquad (8.9)
$$

where the dot denotes differentiation with respect to $\tau = -k_{54}t$, and

$$x_1 \triangleq \frac{-k_{54}}{\delta g_{31} \cdot \delta_{e_{max}}} q$$

$$x_2 \triangleq \frac{k_{54}^2}{\delta g_{31} \cdot \delta_{e_{max}}} \alpha$$

$$x_3 \triangleq \frac{k_{54}^2}{\delta g_{31} \cdot \delta_{e_{max}}} q^{(1)}$$

$$x_4 \triangleq \frac{-k_{54}^3}{\delta g_{31} \cdot \delta_{e_{max}}} \alpha^{(1)}$$

$$x_5 \triangleq \frac{k_{54}^2}{\delta g_{31} \cdot \delta_{e_{max}}} q^{(2)}$$

$$x_6 \triangleq \frac{-k_{54}^3}{\delta g_{31} \cdot \delta_{e_{max}}} \alpha^{(2)}$$

$$u \triangleq \frac{\delta_e}{\delta_{e_{max}}}$$

$$c_1 \triangleq \frac{k_{32}}{k_{54}}$$

$$c_2 \triangleq \frac{-k_{34}}{k_{54}^2}$$

In terms of the normalized variables, the information matrix becomes

$$I_a = \frac{-\delta g_{31}^2 \ \delta_{e_{max}}^2}{Rk_{54}^5} \begin{bmatrix} \int_o^{T'} x_3^2 d\tau & \int_o^{T'} x_3 x_5 d\tau \\ \int_o^{T'} x_3 x_5 d\tau & \int_o^{T'} x_5^2 d\tau \end{bmatrix} \tag{8.10}$$

where T' is the length of the test in normalized units of time. The input performance index is then

$$\phi = \frac{-Rk_{54}^5}{\delta g_{31}^2 \ \delta_{e_{max}}^2} \phi' \tag{8.11}$$

where

$$\phi' = \frac{x_7(T') + x_8(T')}{x_7(T')x_8(T') - x_9^2(T')} \tag{8.12}$$

and

$$\begin{aligned} \dot{x}_7 &= x_5^2 \ , & x_7(0) &= 0 \\ \dot{x}_8 &= x_3^2 \ , & x_8(0) &= 0 \\ \dot{x}_9 &= x_3 x_5 \ , & x_9(0) &= 0 \ . \end{aligned} \tag{8.13}$$

The evaluation of ϕ for a given input requires nine integrations (2 state equations, 4 sensitivity equations, and 3 quadratures) the length of the test.

B.1 Gradient of the Performance Index

To calculate the gradient of the input performance index we must calculate $\partial x_i/\partial t_k$ for $i = 1, 2, \ldots 9$. For $t < t_k$, we have

$$\frac{\partial x_i}{\partial t_k} = 0, \qquad i = 1, 2, \ldots 9 \ . \tag{8.14}$$

At $t = t_k$ we have

$$\frac{\partial x_i}{\partial t_k} = \dot{x}_i\Big|_{t=t_k^-} - \dot{x}_i\Big|_{t=t_k^+} , \qquad i = 1,2,\ldots 9 \tag{8.15}$$

which equals zero except for $i = 1$ which is

$$\frac{\partial x_1}{\partial t_k} = u(t_k^-) - u(t_k^+) = \pm 2 . \tag{8.15'}$$

For $t > t_k$, we must integrate a set of 15 differential equations. The first six equations are given by (8.9) with the values they had at $t = t_k$ as initial conditions. The last nine equations (with initial conditions given above) are:

$$\begin{aligned}
\dot{x}_7 &= -c_1x_7 - c_2x_8 & x_7(t_k) &= \pm 2\\
\dot{x}_8 &= x_7 - x_8 & x_8(t_k) &= 0\\
\dot{x}_9 &= -c_1x_9 - c_2x_{10} + x_7 - x_8 , & x_9(t_k) &= 0\\
\dot{x}_{10} &= x_9 - x_{10} & x_{10}(t_k) &= 0\\
\dot{x}_{11} &= -c_1x_{11} - c_2x_{12} + x_7 & x_{11}(t_k) &= 0\\
\dot{x}_{12} &= x_{11} - x_{12} & x_{12}(t_k) &= 0\\
\dot{x}_{13} &= 2x_5x_{11} & x_{13}(t_k) &= 0\\
\dot{x}_{14} &= 2x_3x_9 & x_{14}(t_k) &= 0\\
\dot{x}_{15} &= x_3 \cdot x_{11} + x_5x_9 & x_{15}(t_k) &= 0 ;
\end{aligned}$$

where x_7 through x_{15} designate $\partial x_1/\partial t_k$ through $\partial x_9/\partial t_k$. The gradient of the input performance index with respect to the kth switch time is then

$$\frac{\partial \phi'}{\partial t_k} = \frac{x_{13}(T') + x_{14}(T')}{D} - \frac{[x_7(T') + x_8(T')]}{D^2}$$

$$\times \left[x_{13}(T')\, x_8(T') + x_7(T') x_{14}(T') - 2x_9(T') x_{15}(T')\right]$$

where $$D = x_7(T')\, x_8(T') - x_9^2(T') .$$

A computer program for the optimal inputs is shown in Appendix A.

C. RESULTS

For one switch (N = 1), a plot of ϕ' was made versus t_1 for various test lengths, namely, $T' = 1, 3, 5,$ and 10 time units. These are plotted in Figs. 8.1a through 8.1d. For the first three cases, there was only one central minimum. For the last case, we see two local minima, the one on the left being the lower. Since the inverted plateau of this latter case is quite long (and the performance therefore rather insensitive to changes in the switch time), we might suspect that only one switch is not a global minimum for $T' = 10$ time units.

For each of the figures 8.1a through 8.1d, there was also a local minimum at $t_1 = 0$. This corresponds to the $N = 0$ case (i.e., no switches). In general, we may say that for the N switch case, there is a local minimum corresponding to the N-1 case. In using the algorithm developed in the previous Chapter, our initial values of the switch times are near the center, so that we converge to a central minimum.

A plot was made of ϕ' for the optimal switch times for $N = 0$ through $N = 3$ and is shown in Fig. 8.2. The lowest value of ϕ' from this curve and the switch times are shown in Fig. 8.3. This, then, is the solution curve. For example, if we wanted to know what the optimal input is for a 10 sec test, we would look under $T' = -k_{54} \cdot 10 = 7.37$ time units. At this test length, $t_1' = 3.00$, $t_2' = 6.18$, and $\phi' = 61$. In other words, the optimal input is full elevator on for 4.07 sec, then full elevator on in opposite direction for 4.31 sec, and then full elevator on in the original direction for 1.62 secs.

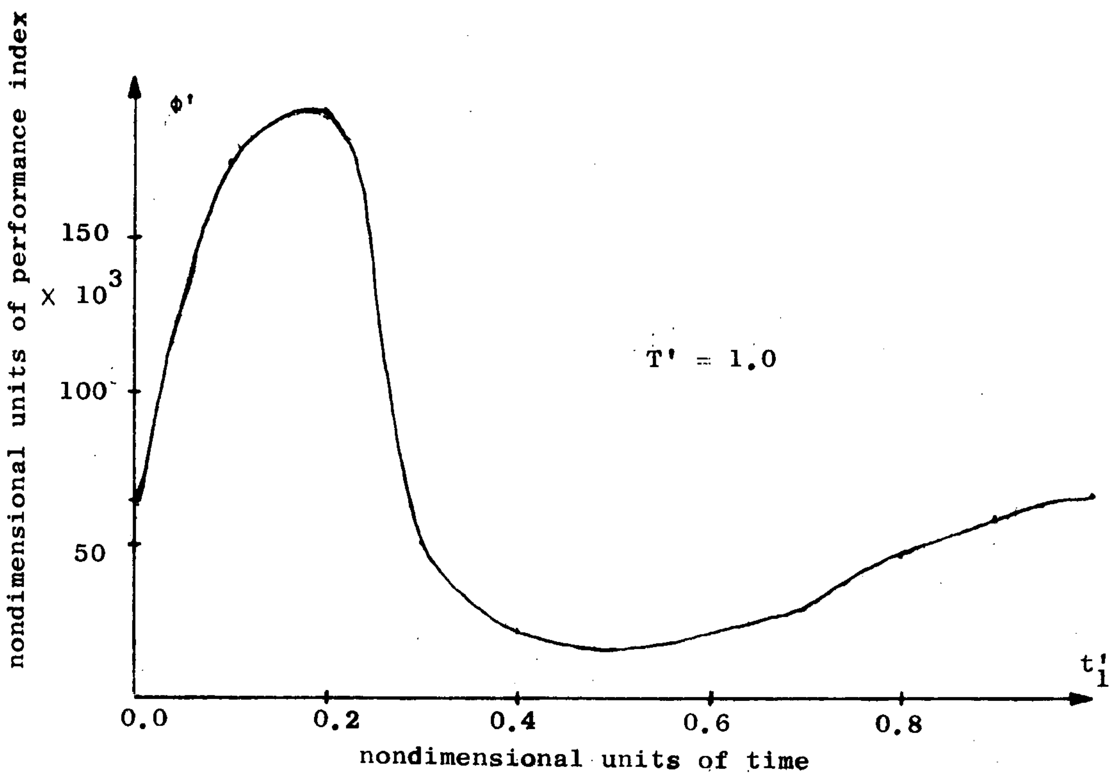


Fig. 8.1a

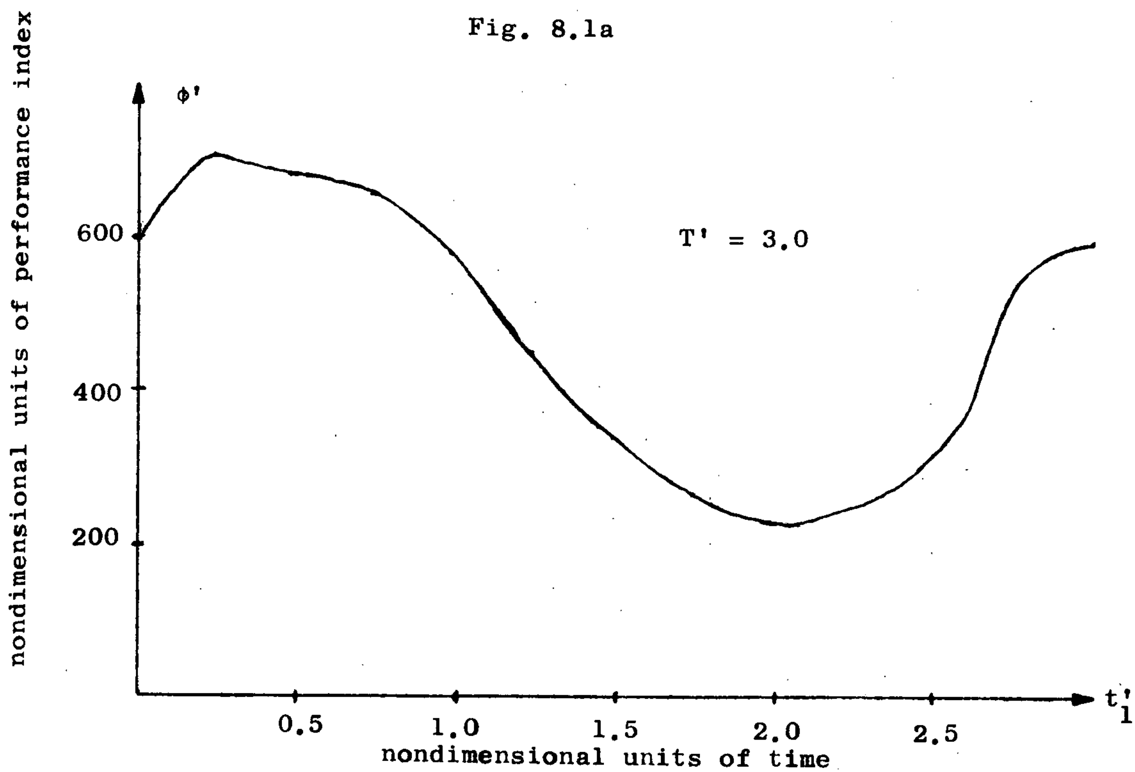


Fig. 8.1b

FIG. 8.1 PERFORMANCE INDEX VS ONE SWITCH TIME.

(Cont)

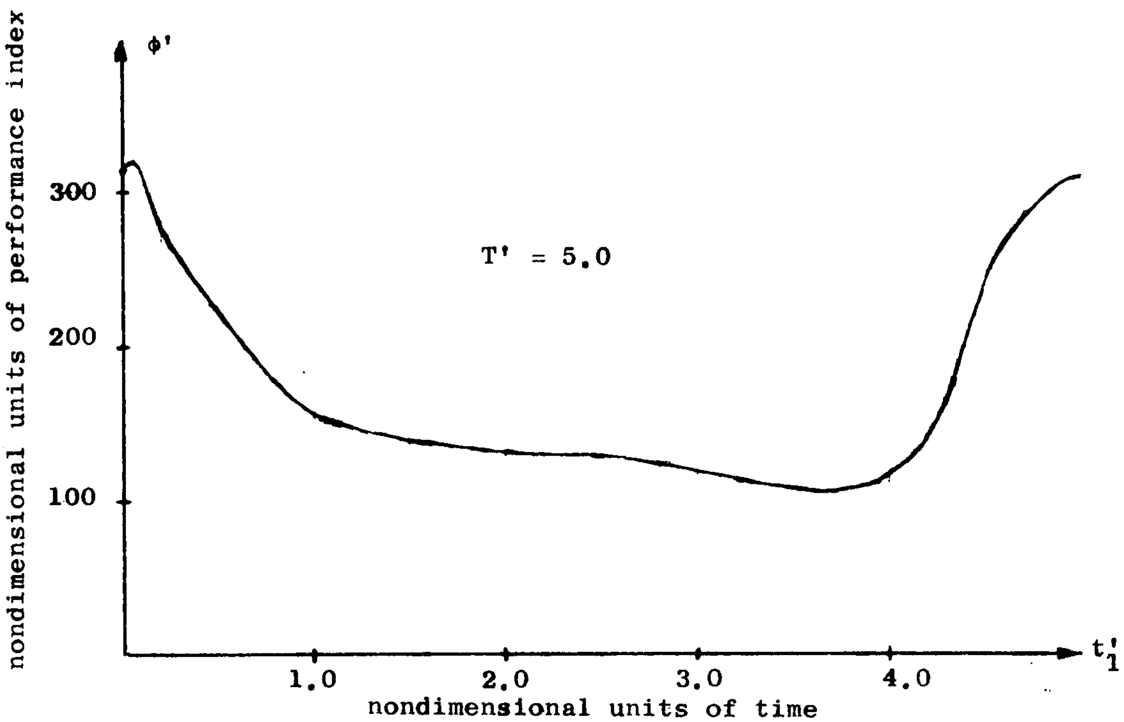


Fig. 8.1c

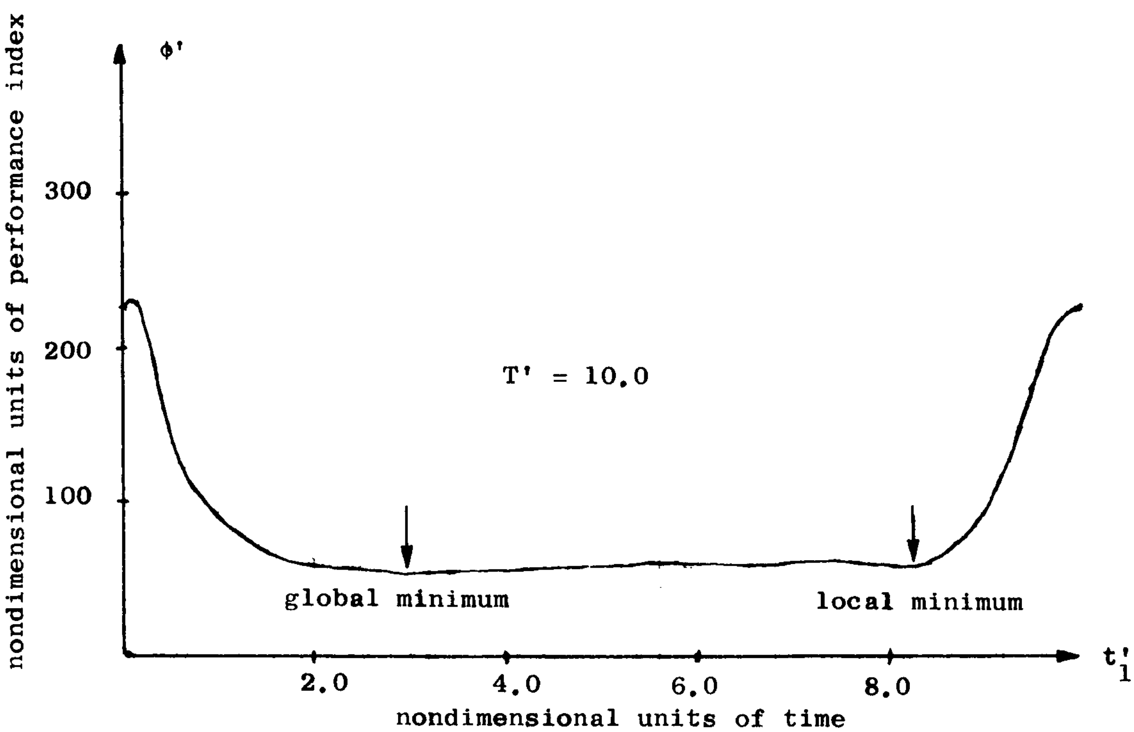


Fig. 8.1d

FIG. 8.1 (Cont) PERFORMANCE INDEX VS ONE SWITCH TIME.

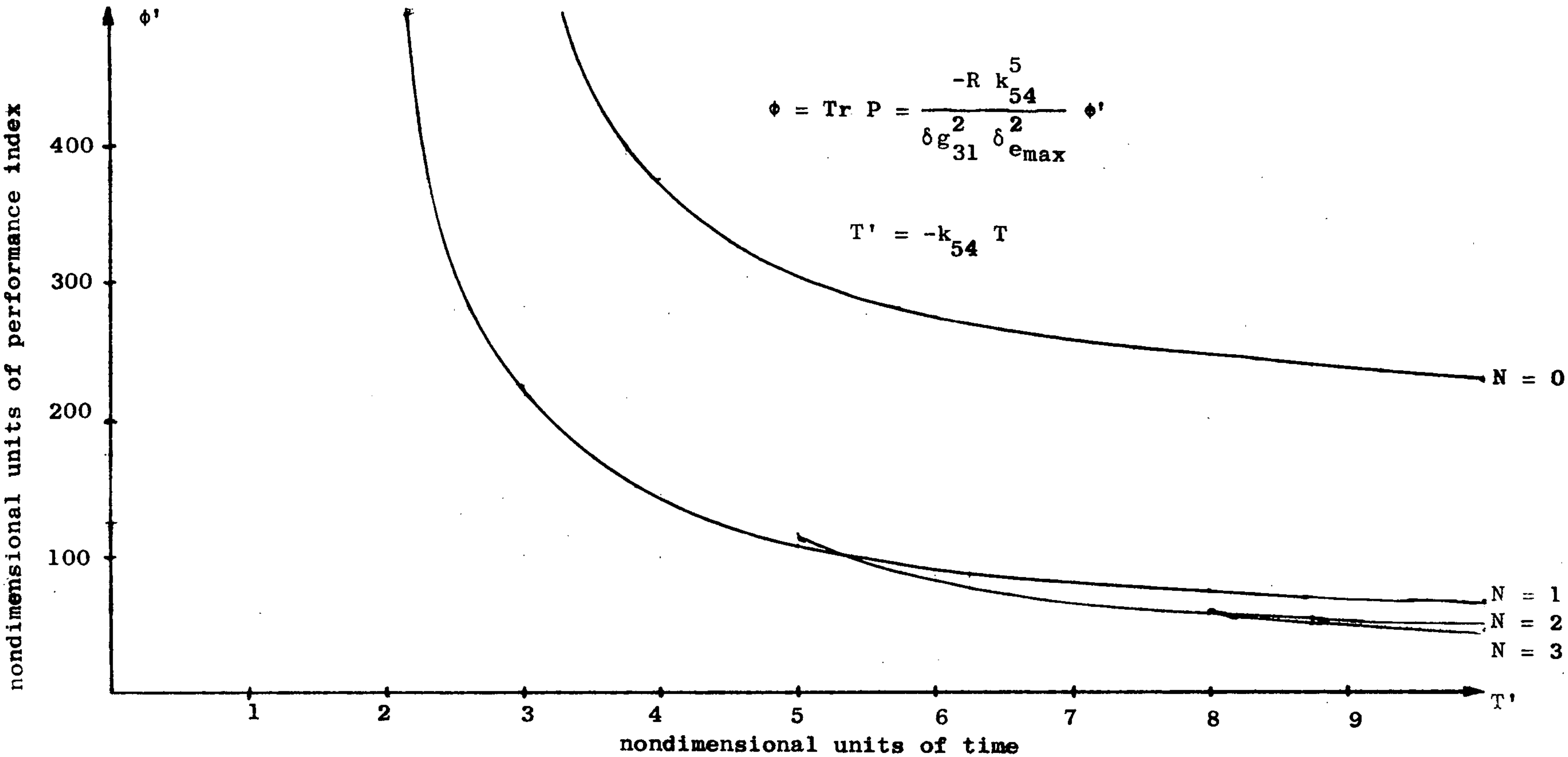


FIG. 8.2 PERFORMANCE INDEX VS LENGTH OF TEST

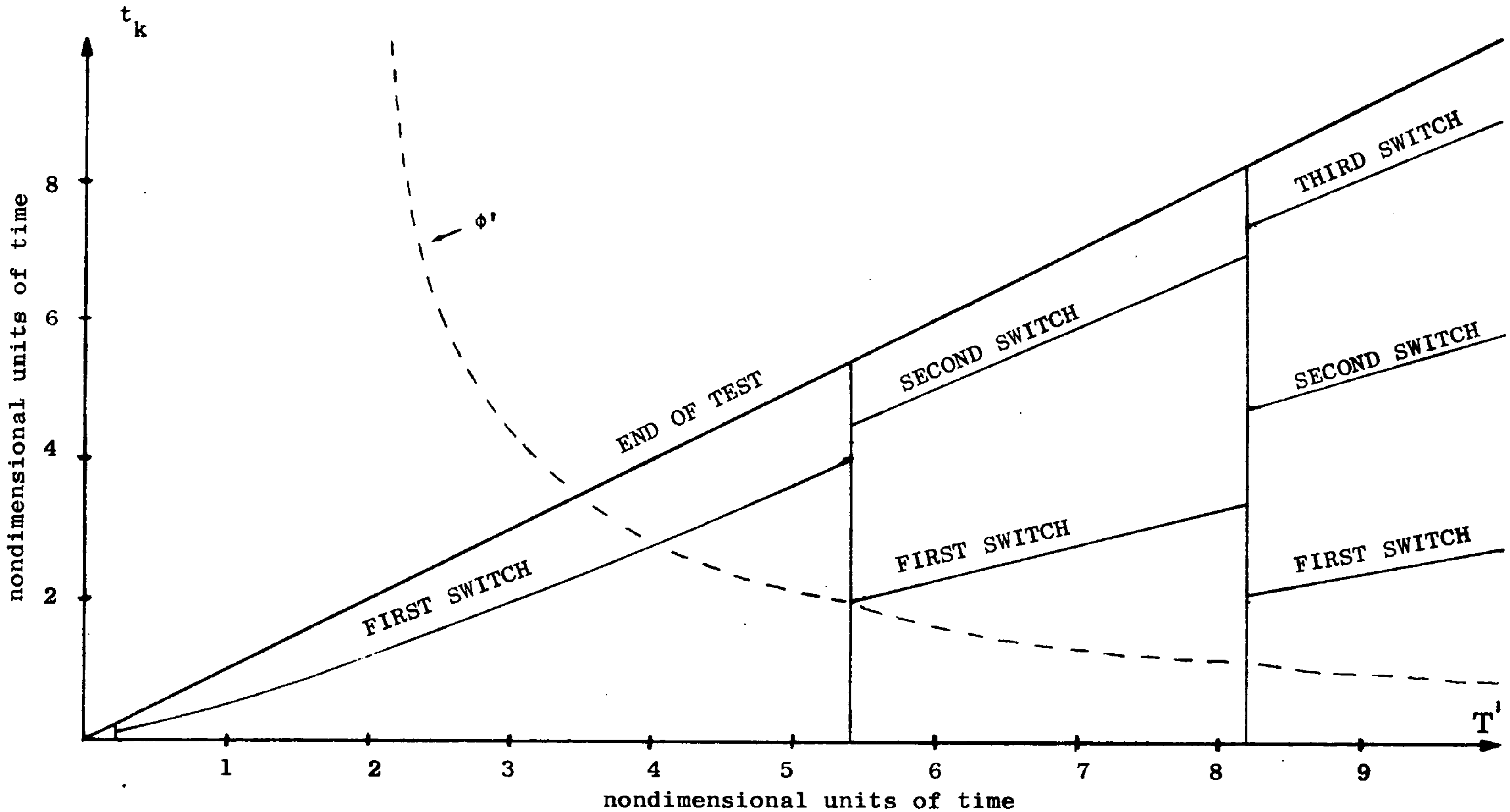


FIG. 8.3. SWITCH TIMES VS LENGTH OF TEST

D. STEADY STATE SOLUTION

For a very long test, we can approximate the repetitive bang-bang inputs with a sine wave. The system and sensitivity equations consist of three second-order systems of the form

$$\begin{aligned} \dot{x}_1 &= -c_1 x_1 - c_2 x_2 + u \\ \dot{x}_2 &= x_1 - x_2 \end{aligned} \tag{8.18}$$

where $x_1 - x_2$ replaces u for the first set of sensitivity equations and x_1 replaces u for the second set of sensitivity equations. The transfer functions are given by

$$\frac{x_1(s)}{u(s)} = \frac{s + 1}{s^2 + s(c_1 + 1) + c_1 + c_2}$$

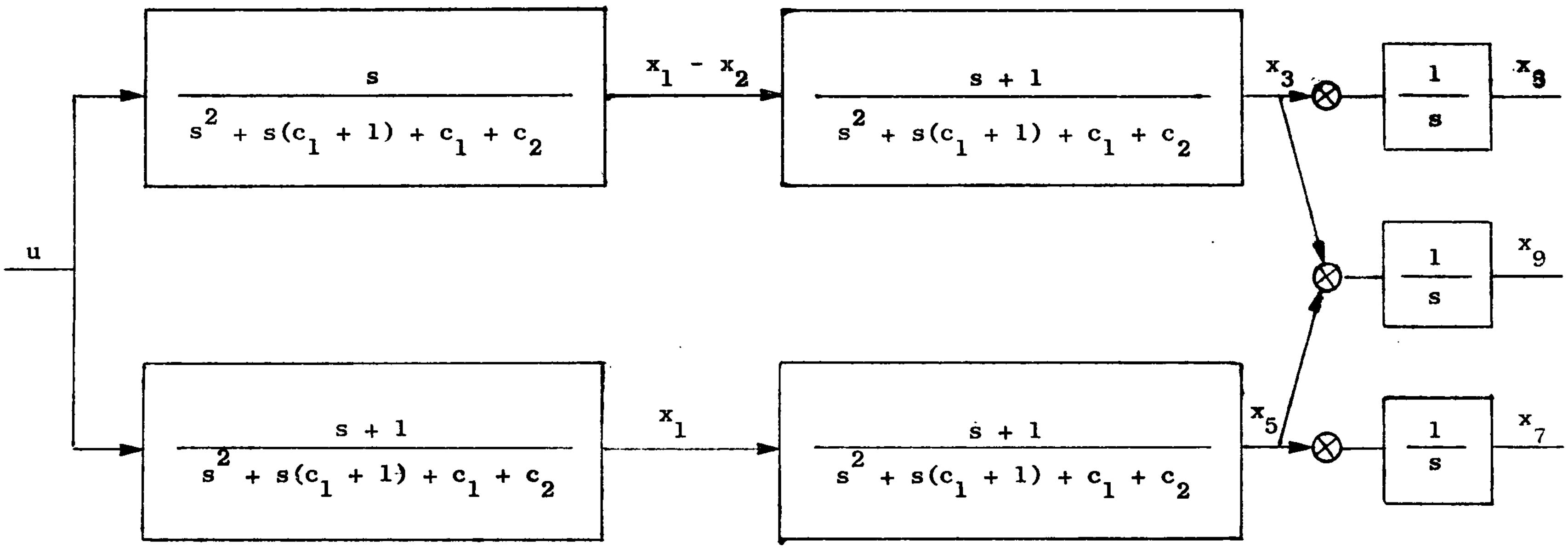
$$\frac{x_2(s)}{u(s)} = \frac{1}{s^2 + s(c_1 + 1) + c_1 + c_2} .$$

A block diagram for the calculation of ϕ is shown in Fig. 8.4. If the input is approximated by $u = 4/\pi \sin \omega t$, then for a long test

$$\begin{aligned} x_7(T') &= \frac{8}{\pi^2} T' M^4 (\omega^2 + 1)^2 \\ x_8(T') &= \frac{8}{\pi^2} T' M^4 \omega^2 (\omega^2 + 1) \\ x_9(T') &= \frac{8}{\pi^2} T' \cos\theta \, M^4 \omega (\omega^2 + 1)^{3/2} \end{aligned} \tag{8.20}$$

where M and θ are defined by

F



$$u = \frac{4}{\pi} \sin \omega t$$

$$x_1 - x_2 = \frac{4}{\pi} M\omega \sin(\omega t+\theta),$$

$$x_1 = \frac{4}{\pi} M(\omega^2+1)^{\frac{1}{2}} \sin \omega t,$$

$$x_3 = \frac{4}{\pi} M^2\omega(\omega^2+1)^{\frac{1}{2}} \sin(\omega t+\theta),$$

$$x_5 = \frac{4}{\pi} M^2(\omega^2+1) \sin \omega t,$$

$$x_8 = \frac{16}{\pi^2} M^4\omega^2(\omega^2+1) \sin^2(\omega t+\theta)$$

$$x_9 = \frac{16}{\pi^2} M^4\omega(\omega^2+1)^{\frac{3}{2}} \sin \omega t \sin(\omega t+\theta)$$

$$x_7 = \frac{16}{\pi^2} M^4(\omega^2+1)^2 \sin^2 \omega t$$

FIG. 8.4 BLOCK DIAGRAM FOR STATE, SENSITIVITY FUNCTIONS, AND ELEMENTS OF THE INFORMATION MATRIX

$$M = \frac{1}{\left[(c_1 + c_2 - \omega^2)^2 + \omega^2(c_1 + 1)^2\right]^{\frac{1}{2}}}$$

$$\sin\theta = \frac{1}{(\omega^2 + 1)^{\frac{1}{2}}} \quad . \qquad (8.21)$$

Substituting (8.20) into (8.12) and simplifying, we have

$$\phi' = \frac{2\omega^2 + 1}{\frac{8}{\pi^2} T' M^4 \omega^2(\omega^2 + 1)(1 - \cos^2\theta)} \quad . \qquad (8.22)$$

Substituting for $M(\omega)$ and θ we then have

$$\phi' = \frac{\pi^2(2\omega^2 + 1)\left[(c_1 + c_2 - \omega^2)^2 + \omega^2(c_1 + 1)^2\right]^2}{8T'\omega^2(\omega^2 + 1)} \quad . \qquad (8.23)$$

For $c_1 + c_2 = 3.185$ and $c_1 + 1 = 3.15$, this has a minimum of $\phi'_{min} = 398/T'$ for $\omega = 1.05$. This corresponds to a switch time interval of 2.99 time units or 4.05 seconds. This is in agreement with the solution curve, Fig. 8.3.

E. SIMULATION

A simulation was run using Denery's combined algorithm to identify $p_1 = M_{\dot{\alpha}}/I_y$ and $p_2 = M_q/I_y$ from measurements of the pitch rate q. The computer program for the simulation is shown in Appendix B.

Recall from Table 8.1 that

$$p_1 + p_2 = k_{32}$$

$$p_1 \cdot k_{54} + m_1 = k_{34} \qquad (8.24)$$

so that if we can estimate k_{32} and k_{34} we can estimate p_1 and p_2 according to

$$\hat{p}_1 = \frac{\hat{k}_{34} - m_1}{k_{54}}$$

$$\hat{p}_2 = \hat{k}_{32} - \hat{p}_1 \, . \tag{8.25}$$

To use Denery's [DE-1] combined algorithm, it is necessary to transform to a canonical form where the unknowns are coefficients of the measured state q. The equations of motion take the following form:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{bmatrix} f_{11} & 1 \\ f_{21} & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} g_{11} \\ g_{21} \end{pmatrix} \delta_e \; ; \; \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} = 0 \tag{8.26}$$

$$q = [1 \;\; 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{8.27}$$

where $x_1 = q$ and $x_2 = k_1 q + k_2 \alpha$ so that

$$\dot{q} = (f_{11} + k_1) q + k_2 \alpha + g_{11} \cdot \delta_e$$

$$\dot{\alpha} = \frac{1}{k_2} \left[f_{21} - k_1 (f_{11} + k_1) \right] q - k_1 \alpha + \frac{1}{k_2} (g_{21} - k_1 g_{11}) \delta_e \, . \tag{8.28}$$

By matching the coefficients in equations 8.28 with the first two equations in set (8.8) we have

$$g_{11} = \delta g_{31} \text{ (known)}$$

$$k_2 = k_{34}$$

$$f_{11} = k_{32} - k_1 = k_{32} + k_{54} \tag{8.29}$$

$$k_1 = -k_{54} \text{ (known)}$$

$$f_{21} = k_2 + k_1 \cdot k_{32} = k_{34} - k_{54} \cdot k_{32}$$
$$g_{21} = k_1 \cdot \delta g_{11} = -k_{54} \cdot \delta g_{31} \quad \text{(known)} \; . \qquad (8.29) \text{ cont.}$$

If we can identify f_{11} and f_{21} from Denery's algorithm, then we can calculate k_{32} and k_{34} from

$$\hat{k}_{32} = \hat{f}_{11} - k_{54}$$
$$\hat{k}_{34} = \hat{k}_2 = \hat{f}_{21} + k_{54} \cdot \hat{k}_{32} \; . \qquad (8.30)$$

Notice that we cannot identify all six stability derivatives $(M_{\dot{\alpha}}/I_y$, M_q/I_y, z_α/mu_o, z_{δ_e}/mu_o, M_α/I_y, $M_{\delta_e}/I_y)$ from the five coefficients $(k_{32}$, k_{34}, δg_{31}, k_{54}, $\delta g_{51})$ and with a scalar measurement we cannot identify the above five coefficients from the four canonical coefficients $(f_{11}$, f_{21}, g_{11}, $g_{21})$. Since we are only trying to identify two stability derivatives, the scalar measurement is satisfactory.

For simulation purposes we use values for the stability derivatives calculated from the five coefficients identified in Denery's 17-second test. However, one other stability derivative such as $M_{\dot{\alpha}}/I_y$ is needed or we may make an assumption such as $M_{\dot{\alpha}} = M_q$. The numerical values for this simulation were shown in Table 8.1.

Now applying Denery's algorithm to the second-order system (8.26), we have

$$F_n = \begin{bmatrix} f_{11}^n & 1 \\ f_{21}^n & 0 \end{bmatrix} \qquad H_n = [1 \quad 0] \qquad (8.31)$$

where F_n and H_n are given by

$$F_n = F - DH \tag{8.32}$$

$$H_n = (I - L)H \tag{8.33}$$

so that for this example

$$L = 0, \quad \delta G = 0, \quad \delta x_o = 0, \quad D = \begin{pmatrix} D_{11} \\ D_{21} \end{pmatrix}. \tag{8.33}$$

Estimates of f_{11} and f_{21} are given by

$$\begin{aligned} \hat{f}_{11} &= f^n_{11} + D_{11} \\ \hat{f}_{21} &= f^n_{21} + D_{21} . \end{aligned} \tag{8.34}$$

The simulated measurement z, is given by $z = x_1 + v$ where

$$\begin{aligned} \dot{x}_1 &= f_{11}x_1 + x_2 + g_{11} \cdot \delta_e , \qquad x_1(0) = 0 \\ \dot{x}_2 &= f_{21}x_1 \qquad\quad + g_{21} \cdot \delta_e , \qquad x_2(0) = 0 . \end{aligned} \tag{8.35}$$

The nominal output is given by $z_n = x_{n1}$ where

$$\begin{pmatrix} \dot{x}_{n1} \\ \dot{x}_{n2} \end{pmatrix} = \begin{matrix} f^n_{11} x_{n1} + x_{n2} + g_{11} \cdot \delta_e , & x_{n1}(0) = 0 \\ f^n_{21} x_{n1} \qquad\quad + g_{21} \cdot \delta_e , & x_{n2}(0) = 0 . \end{matrix} \tag{8.36}$$

The sensitivity equations for D_{11} and D_{21} are given by

$$\begin{aligned} \left(\frac{\partial \dot{x}_{n1}}{\partial D_{11}}\right) &= f^n_{11}\left(\frac{\partial x_{n1}}{\partial D_{11}}\right) + \left(\frac{\partial x_{n2}}{\partial D_{11}}\right) + z \quad (\text{or } z_n) \\ \left(\frac{\partial \dot{x}_{n2}}{\partial D_{11}}\right) &= f^n_{21}\left(\frac{\partial x_{n1}}{\partial D_{11}}\right) \end{aligned} \tag{8.37}$$

and

$$\left(\frac{\partial \dot{x}_{n1}}{\partial D_{21}}\right) = f_{11}^{n}\left(\frac{\partial x_{n1}}{\partial D_{21}}\right) + \left(\frac{\partial x_{n2}}{\partial D_{21}}\right)$$

$$\left(\frac{\partial \dot{x}_{n2}}{\partial D_{21}}\right) = f_{21}^{n}\left(\frac{\partial x_{n1}}{\partial D_{21}}\right) + z \quad (\text{or } z_n) \; . \tag{8.38}$$

Estimates of D_{11} and D_{21} are given by

$$\begin{bmatrix} \hat{D}_{11} \\ \hat{D}_{21} \end{bmatrix} = r \begin{bmatrix} \int_o^T \left(\frac{\partial x_{n1}}{\partial D_{11}}\right)^2 dt & \int_o^T \left(\frac{\partial x_{n1}}{\partial D_{11}}\right)\left(\frac{\partial x_{n1}}{\partial D_{21}}\right) dt \\ \int_o^T \left(\frac{\partial x_{n1}}{\partial D_{11}}\right)\left(\frac{\partial x_{n1}}{\partial D_{21}}\right) dt & \int_o^T \left(\frac{\partial x_{n1}}{\partial D_{21}}\right)^2 dt \end{bmatrix}^{-1} \tag{8.39}$$

$$\times \frac{1}{r} \begin{bmatrix} \int_o^T \frac{\partial x_{n1}}{\partial D_{11}} (z - z_n)\, dt \\ \int_o^T \frac{\partial x_{n1}}{\partial D_{21}} (z - z_n)\, dt \end{bmatrix} .$$

By combining the linear transformations

$$\begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \end{pmatrix} = \begin{bmatrix} \frac{1}{k_{54}} & 0 \\ -\frac{1}{k_{54}} & 1 \end{bmatrix} \begin{pmatrix} \hat{k}_{34} \\ \hat{k}_{32} \end{pmatrix} + \begin{pmatrix} -\frac{m_1}{k_{54}} \\ \frac{m_1}{k_{54}} \end{pmatrix} \tag{8.40}$$

and

$$\begin{pmatrix} \hat{k}_{32} \\ \hat{k}_{34} \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ k_{54} & 1 \end{bmatrix} \begin{pmatrix} \hat{f}_{11} \\ \hat{f}_{21} \end{pmatrix} + \begin{pmatrix} -k_{54} \\ -k_{54}^2 \end{pmatrix} , \tag{8.41}$$

we find that

$$\begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \end{pmatrix} = \begin{bmatrix} 1 & \frac{1}{k_{54}} \\ 0 & -\frac{1}{k_{54}} \end{bmatrix} \begin{pmatrix} \hat{f}_{11} \\ \hat{f}_{21} \end{pmatrix} + \begin{pmatrix} -k_{54} - \frac{m_1}{k_{54}} \\ \frac{m_1}{k_{54}} \end{pmatrix} . \tag{8.42}$$

The covariance in our estimates of the parameters p_1 and p_2 is given in terms of the covariance of f_{11} and f_{21} by

$$P = E(p - \bar{p})(p - \bar{p})^T \tag{8.43}$$

$$= \begin{bmatrix} P_{D11} + \frac{2}{k_{54}} P_{D12} + \frac{1}{k_{54}^2} P_{D22} & -\frac{1}{k_{54}} P_{D12} - \frac{1}{k_{54}^2} P_{D22} \\ -\frac{1}{k_{54}} P_{D12} - \frac{1}{k_{54}^2} P_{D22} & \frac{1}{k_{54}^2} P_{D22} \end{bmatrix} .$$

For a four-second test, $T' = -k_{54} \cdot 4 \approx 3.00$. From Fig. 8.3, we see that for $T' = 3$, $N = 1$ is optimal with $t_1' = 2.04$ and $\phi' = 228$. In this case the normalized covariance for D_{11} and D_{21} (or f_{11} and f_{21}) is

$$P_D' = \begin{bmatrix} 150 & -80 \\ -80 & 78 \end{bmatrix} . \tag{8.44}$$

The predicted covariance for D_{11} and D_{21} is then

$$P_D = \frac{-Rk_{54}^5}{\delta g_{31}^2 \cdot \delta_{e_{max}}} \bar{P} = R \begin{bmatrix} 297 & -159 \\ -159 & 156 \end{bmatrix} . \tag{8.45}$$

Substituting values in (8.45) with $R = \sigma^2 \Delta t = (0.1)^2(0.01) = 10^{-4}$ into (8.43), the predicted covariance matrix for p_1 and p_2 is

$$P = \begin{bmatrix} 0.0297 & -0.0159 \\ -0.0159 & 0.0156 \end{bmatrix} .$$

In the simulation, Denery's algorithm was applied to 20 sets of data and the results are summarized in Table 8.2. Except as noted in the first column, all of the tests had a bang-bang input with a switch at 2.72 sec, a standard deviation in measurements of 0.1 rad/sec, and an initial guess of p_1 and p_2 of -0.60 and -0.80 respectively. The average number of iterations for the 20 tests is shown for the equation error and the output error portions of Denery's algorithm in columns 2 and 3. The resultant covariance of the estimates is shown in columns 4 through 6.

From Table 8.2 we can make the following conclusions: (1) With an optimal input, Denery's algorithm converges faster and to a more accurate estimate than with a non-optimal input. (2) The predicted covariance given by the inverse of the information matrix is very close to that calculated in the simulation. (3) For large errors in the initial estimates of the unknown parameters, the equation error portion takes more iterations to converge; but, the number of output error iterations remains the same. (4) An indication of the final accuracy in our estimates is provided by the number of iterations needed for the output error portion of the algorithm to converge. In a sense, then, the bias from the equation error portion serves a useful purpose.

A computer listing of this simulation is shown in Appendix B.

Table 8.2

RESULTS OF DENERY'S IDENTIFICATION ALGORITHM APPLIED TO 20 TESTS. Except as noted in the first column, each set of tests had a bang-bang input with a switch at $t_1 = 2.72$ sec, a standard deviation in measurements of $\sigma_r = 0.1$ rad/sec, and an initial guess of the parameters of $p_1 = -0.60$ and $p_2 = -0.80$.

Condition	Equation Error Iterations	Output Error Iterations	Covariance Matrix		
			P11	P12	P22
Optimal, $t_1 = 2.72$	3.45	2.20	0.026946	-0.015152	0.015005
$u = 0.2 \sin 1.57t$	3.85	3.00	0.073325	-0.058342	0.054742
$t_1 = 2.65$	3.45	2.25	0.026368	-0.015742	0.015737
$t_1 = 2.79$	3.45	2.75	0.027971	-0.014758	0.014366
Initial Condition, p_1 and $p_2 = -10.0$	5.00	2.20	0.026946	-0.015152	0.015005
$\sigma_r = 0.01$	2.95	1.30	0.000260	-0.000154	0.000154

Chapter IX

OPTIMAL INPUTS FOR THE IDENTIFICATION OF THE LATERAL DYNAMIC STABILITY DERIVATIVES

A. PROBLEM FORMULATION

Approximate lateral equations of motion for a conventional airplane are*

$$\dot{\beta} + r = \frac{Y_\beta}{mV}\beta + \frac{g}{V}\phi$$

$$\dot{r} + \frac{I_{xz}}{I_{zz}}\dot{p} = \frac{n_\beta}{I_{zz}}\beta + \frac{n_r}{I_{zz}}r + \frac{n_p}{I_{zz}}p + \frac{n_{\delta_r}}{I_{zz}}\delta_r$$

$$\dot{p} + \frac{I_{xz}}{I_{xx}}\dot{r} = \frac{\ell_\beta}{I_{xx}}\beta + \frac{\ell_r}{I_{xx}}r + \frac{\ell_p}{I_{xx}}p + \frac{\ell_{\delta_a}}{I_{xx}}\delta_a \qquad (9.1)$$

$$\dot{\phi} = p$$

$$\dot{\psi} = r$$

where

β = sideslip angle

r = yaw angular velocity

p = roll angular velocity

---

* The equations and numerical values used for these computations were taken from Bryson and Ho [BRY-1, p. 173].

ϕ = roll angle

ψ = yaw angle

δ_r = rudder deflection

δ_a = aileron deflection .

We wish to identify the four dynamic stability derivatives n_r, n_p, ℓ_r, and ℓ_p, assuming that the other stability derivatives and the two control derivatives are known from wind tunnel tests. These four dynamic stability derivatives depend upon motion of the aircraft and may be difficult to determine from wind tunnel tests. Let us identify the parameters in the normalized form:

$$p_1 = \frac{n_r}{I_{zz}} , \quad p_2 = \frac{n_p}{I_{zz}} , \quad p_3 = \frac{\ell_r}{I_{xx}} , \quad \text{and} \quad p_4 = \frac{\ell_p}{I_{xx}} .$$

For this example, let us assume that the only output measurements are yaw rate r and roll rate p, each corrupted by uncorrelated white gaussian noises of density R (a scalar). Our problem is to determine the optimal inputs δ_r and δ_a for the identification test.

B. INPUT CRITERION

The identification performance index is

$$J = \frac{1}{2R} \int_o^T (z_1 - r)^2 + (z_2 - p)^2 \, dt \tag{9.2}$$

so that the i,jth element of the information matrix is then

$$I_{ij} = \frac{1}{R} \int_o^T \left(\frac{\partial r}{\partial p_i}\right)\left(\frac{\partial r}{\partial p_j}\right) + \left(\frac{\partial p}{\partial p_i}\right)\left(\frac{\partial p}{\partial p_j}\right) dt \tag{9.3}$$

which is a quadrature of products of the sensitivity functions. As an input performance index, let us choose

$$\phi = \text{Tr } I_a^{-1} . \quad (9.4)$$

The four sets of sensitivity equations for p_1, p_2, p_3 and p_4 are the same as (9.1) except that the inputs are

$$\begin{bmatrix} 0 \\ r \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ p \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ r \\ 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 0 \\ p \\ 0 \\ 0 \end{bmatrix} \text{ instead of } \begin{bmatrix} 0 \\ \frac{n_r}{I_{zz}} \delta_r \\ \frac{\ell_{\delta_a}}{I_{zz}} \delta_a \\ 0 \\ 0 \end{bmatrix}$$

and the "states" are the sensitivity functions

$$\frac{\partial\beta}{\partial p_i}, \frac{\partial r}{\partial p_i}, \frac{\partial p}{\partial p_i}, \frac{\partial\phi}{\partial p_i}, \text{ and } \frac{\partial\psi}{\partial p_i}$$

where $i = 1, 2, 3, 4$. The last equation in each set, ψ and $\partial\psi/\partial p_i$, $i = 1, 2, 3, 4$ is uncoupled from the other equations and may be dropped since there is no state constraint on ψ and we are not using measurements of ψ. The system equations may also be written in the form

$$\begin{aligned} \dot{\beta} &= c_1\beta - r + c_2\phi \\ \dot{r} &= c_3\beta + c_4 r + c_5 p + c_6\delta_r + c_7\delta_a \\ \dot{p} &= c_8\beta + c_9 r + c_{10} p + c_{11}\delta_r + c_{12}\delta_a \\ \dot{\phi} &= p \end{aligned} \quad (9.5)$$

where

$$
\begin{aligned}
c_1 &= \frac{Y_\beta}{mV} & c_8 &= \frac{\left(\frac{\ell_\beta}{I_{xx}} - \frac{I_{xz}}{I_{xx}}\frac{n_\beta}{I_{zz}}\right)}{D} \\
c_2 &= \frac{g}{V} & & \\
c_3 &= \frac{\left(\frac{n_\beta}{I_{zz}} - \frac{I_{xz}}{I_{zz}}\frac{\ell_\beta}{I_{xx}}\right)}{D} & c_9 &= \frac{\left(\frac{\ell_r}{I_{xx}} - \frac{I_{xz}}{I_{xx}}\frac{n_r}{I_{zz}}\right)}{D} \\
c_4 &= \frac{\frac{n_r}{I_{zz}} - \frac{I_{xz}}{I_{zz}}\frac{\ell_r}{I_{xx}}}{D} & c_{10} &= \frac{\left(\frac{\ell_p}{I_{xx}} - \frac{I_{xz}}{I_{xx}}\frac{n_p}{I_{zz}}\right)}{D} \\
c_5 &= \frac{\left(\frac{n_p}{I_{zz}} - \frac{I_{xz}}{I_{zz}}\frac{\ell_p}{I_{xx}}\right)}{D} & c_{11} &= \frac{-\frac{I_{xz}}{I_{xx}}\frac{n_{\delta r}}{I_{zz}}}{D} \\
c_6 &= \frac{n_{\delta r}}{I_{zz}D} & c_{12} &= \frac{\ell_{\delta a}}{I_{xx}D} \\
c_7 &= \frac{-\frac{I_{xz}}{I_{zz}}\frac{\ell_{\delta a}}{I_{xx}}}{D} & D &= 1 - \frac{I_{xz}I_{xz}}{I_{xx}I_{zz}} .
\end{aligned}
\tag{9.6}
$$

The sensitivity equations for p_1, p_2, p_3, and p_4 are of the same form as (9.5) with the following modifications:

(1) For p_1 substitute r for $\frac{n_{\delta r}}{I_{zz}} \delta_r$ and set $\delta_a = 0$.

(2) For p_2 substitute p for $\frac{n_{\delta r}}{I_{zz}} \delta_r$ and set $\delta_a = 0$.

(3) For p_3 substitute r for $\frac{n_{\delta a}}{I_{xx}} \delta_a$ and set $\delta_r = 0$.

(4) For p_4 substitute p for $\frac{\ell_{\delta a}}{I_{xx}} \delta_a$ and set $\delta_r = 0$.

Evaluating the performance index requires 30 integrations (4 state equations, 16 sensitivity equations, and 10 information matrix quadratures) over the interval from 0 to T.

C. GRADIENT OF THE PERFORMANCE INDEX

The partial derivative of the performance index with respect to the kth switch time t_k is given by

$$\frac{\partial\phi}{\partial t_k} = -\mathrm{Tr} I_a^{-1} \frac{\partial I_a}{\partial t_k} I_a^{-1} . \tag{9.7}$$

The elements of $\partial I_a/\partial t_k$ are found by integrating product terms involving x_i and $\partial x_i/\partial t_k$, $i = 1, 2, \ldots 20$. The differential equations for $\partial x_i/\partial t_k$, $i = 1, 2, \ldots 20$ are the same as those for x_i, $i = 1, 2, \ldots 20$, except for the elimination of the inputs δ_r and δ_a. They are integrated forward in time from t_k to T with initial conditions given by

$$\frac{\partial x_i}{\partial t_k}(t_k) = \dot{x}_i\Big|_{t=t_k^-} - \dot{x}_i\Big|_{t=t_k^+} . \tag{9.8}$$

Evaluating the partial derivative of the performance index with respect to the kth switch time requires 50 integrations (20 state and sensitivity equations, 20 equations with respect to t_k, and 10 quadratures for the elements of $\partial I_a/\partial t_k$) over the interval from t_k to T .

With more than one input, the assignment of switch times for each individual input becomes a little more complicated. For this problem the first input δ_r, has N_1 switches at times $t_{1,1}$, $t_{1,2}, \ldots t_{1,N_1}$, and the second input δ_a has N_2 switches at times $t_{2,1}$, $t_{2,2}$, ... t_{2,N_2}. There are a total of N switches at t_1, t_2, ... t_N where $N = N_1 + N_2$. Figure 9.1 shows a possible switch assignment for the case $N_1 = 2$, and $N_2 = 3$. Since the individual switch times are incremented by different amounts, this assignment can change with each iteration.

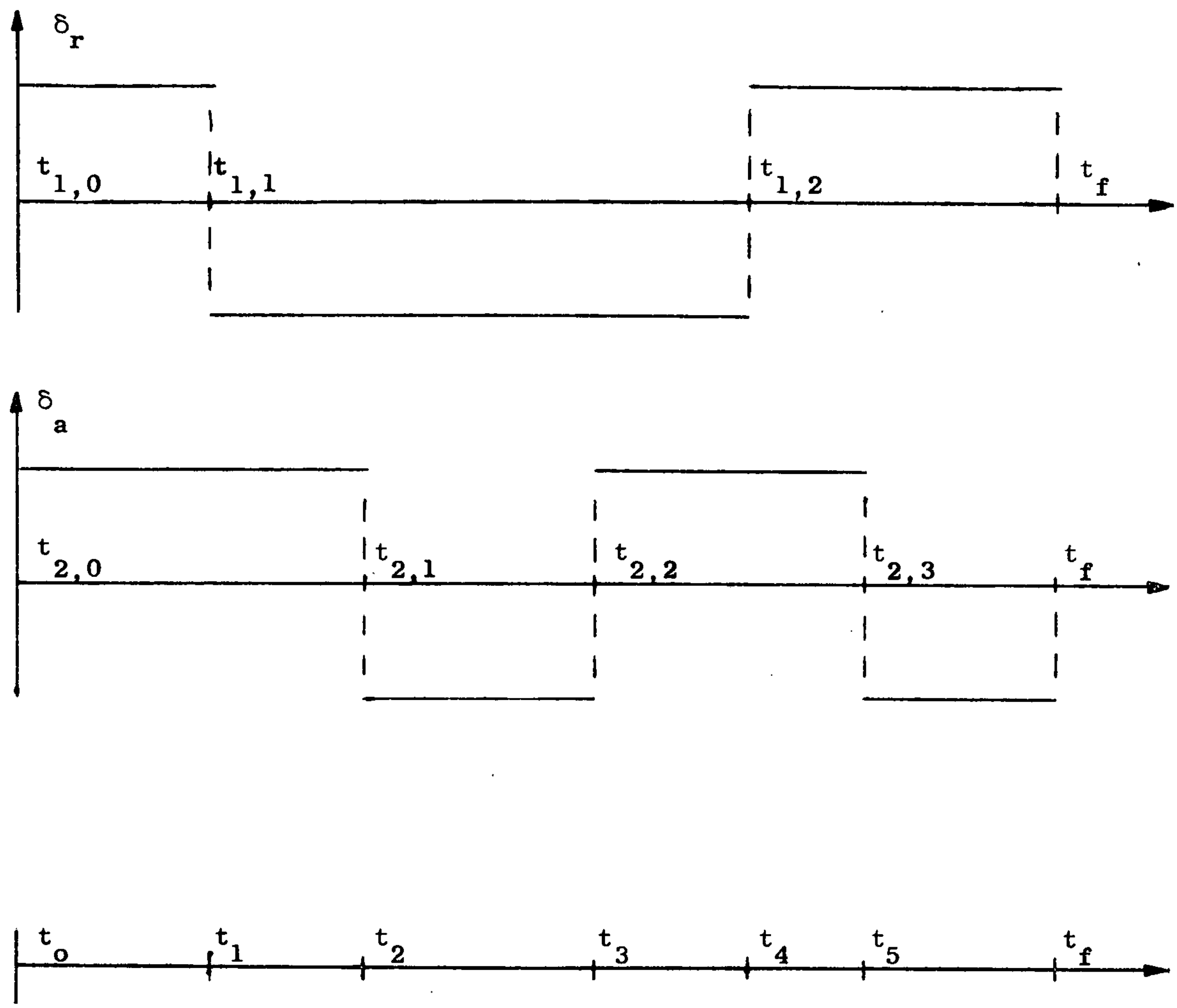


FIG. 9.1 POSSIBLE SWITCH TIME ASSIGNMENTS

D. RESULTS

The conjugate gradient search routine for the optimal switch times is similar to the Chapter VIII Example implemented in Appendix A.

To insure a global minimum, we proceed as before by plotting the optimal performance index for a number of cases that depend upon the number of switches. For the scalar input case, we had one case for N switches. With two inputs, however, we have $2(N+1)$ possible cases for N switches.

For $N = 0$ (no switches) we have two cases: $N_1 = 0$, $N_2 = 0$ (no switches for either input), and the two inputs either start (1) with the same sign, or (2) with opposite signs (i.e., in-phase or out-of-phase, $P = \pm 1$).

For $N = 1$, we have the four cases: (1) $N_1 = 1$, $N_2 = 0$, $P = +1$ (one switch for input δ_r, two inputs initially same sign); (2) $N_1 = 1$, $N_2 = 0$, $P = -1$ (one switch for input δ_r, two inputs initially the opposite sign); (3) $N_1 = 0$, $N_2 = 1$, $P = 1$ (one switch for input δ_a, two inputs initially the same sign); (4) $N_1 = 0$, $N_2 = 1$, $P = -1$ (one switch for input δ_a, two inputs different signs).

Each of these cases is an optimization problem with respect to one parameter. Figure 9.2a to 9.2d show the performance index ϕ versus the parameter of interest for a test length of five seconds. The end values of the performance index correspond to an $N = 0$ case. The performance index versus the length of the test for an optimal input is shown in Fig. 9.3 for each of the six cases of $N = 0$ and $N = 1$. Each case is specified by the triplet (N_1, N_2, P).

For $N = 2$ there are six possible cases, namely: • (2, 0, 1), • (2, 0, -1), • (1, 1, 1), • (1, 1, -1), • (0, 2, 1), • (0, 2, -1). Each of these cases involves an optimization problem with respect to two parameters. Values of the performance index are shown on a grid of the two parameters of interest in Figs. 9.4a to 9.4f for a five second test. Except for the (1, 1, -1) case of Fig. 9.4d, each of these cases has its minimum at a minimum of an $N = 1$ case.

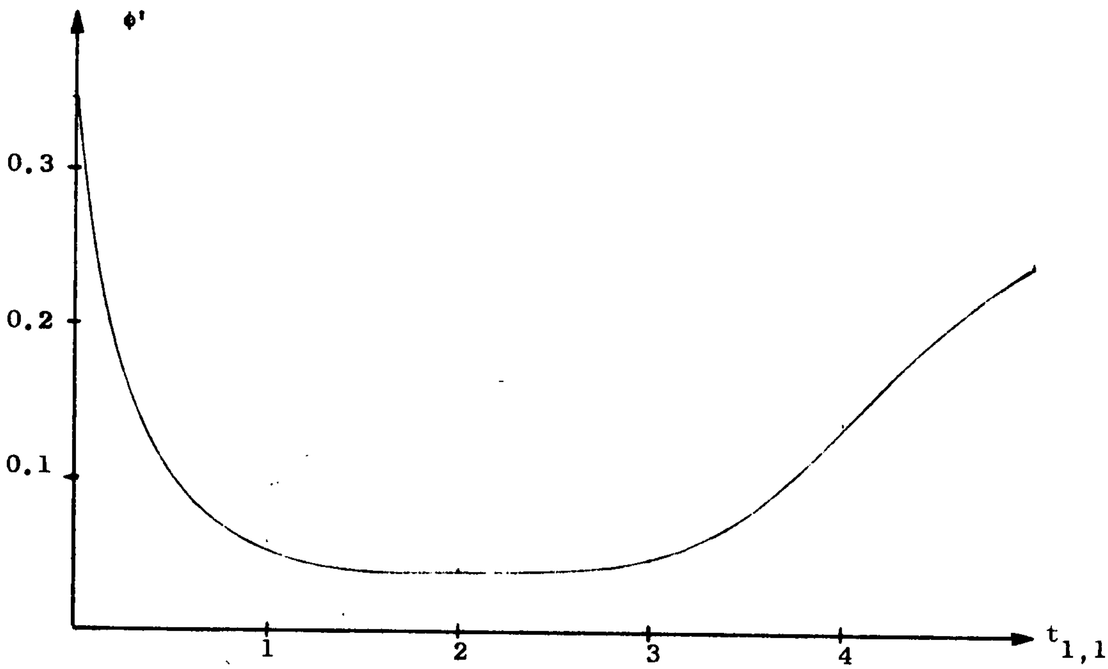


Fig. 9.2a: Case 1: $N_1 = 1$, $N_2 = 0$, $P = +1$

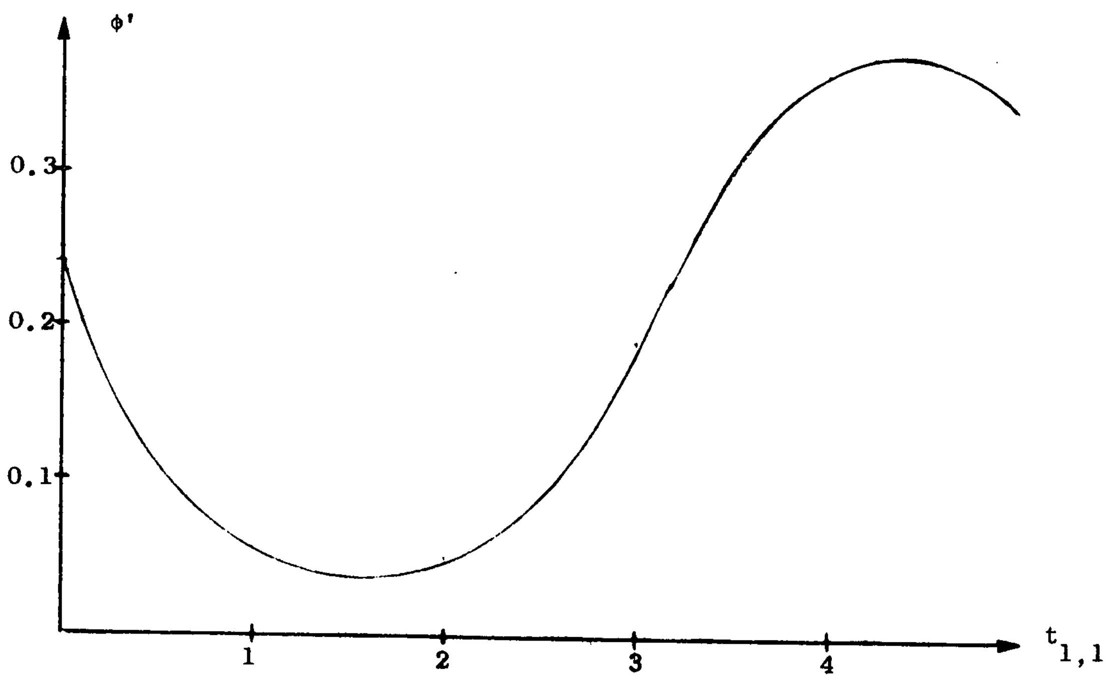


Fig. 9.2b: Case 2: $N_1 = 1$, $N_2 = 0$, $P = -1$

(Cont)

FIG. 9.2 PERFORMANCE INDEX VS ONE SWITCH TIME

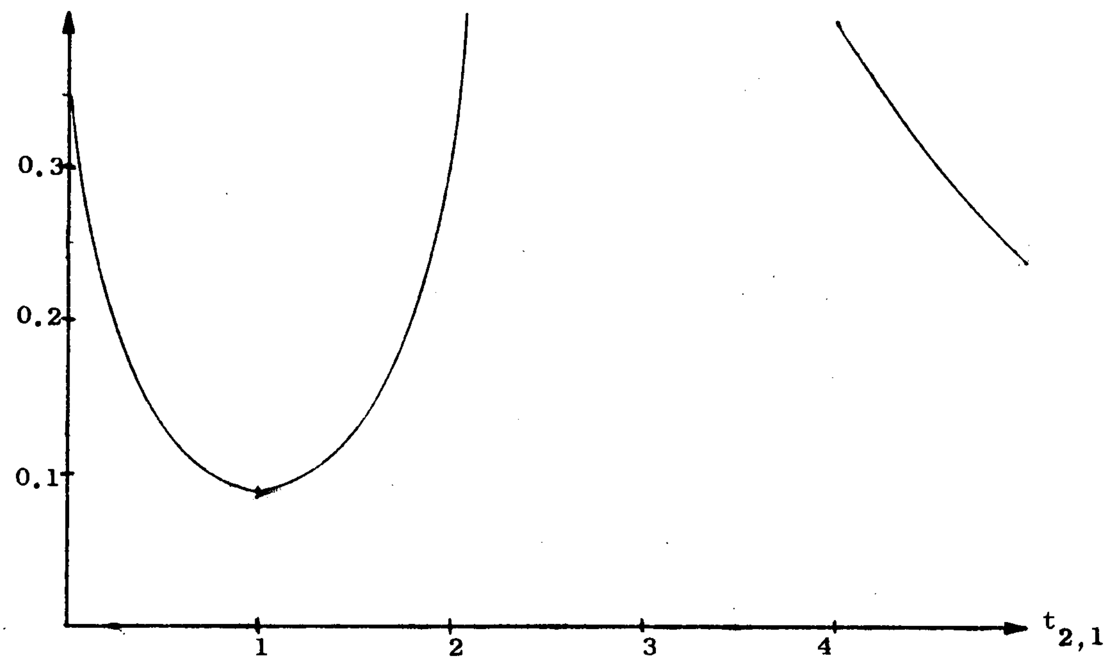


Fig. 9.2c: Case 3: $N_1 = 0$, $N_2 = 1$, $P = +1$

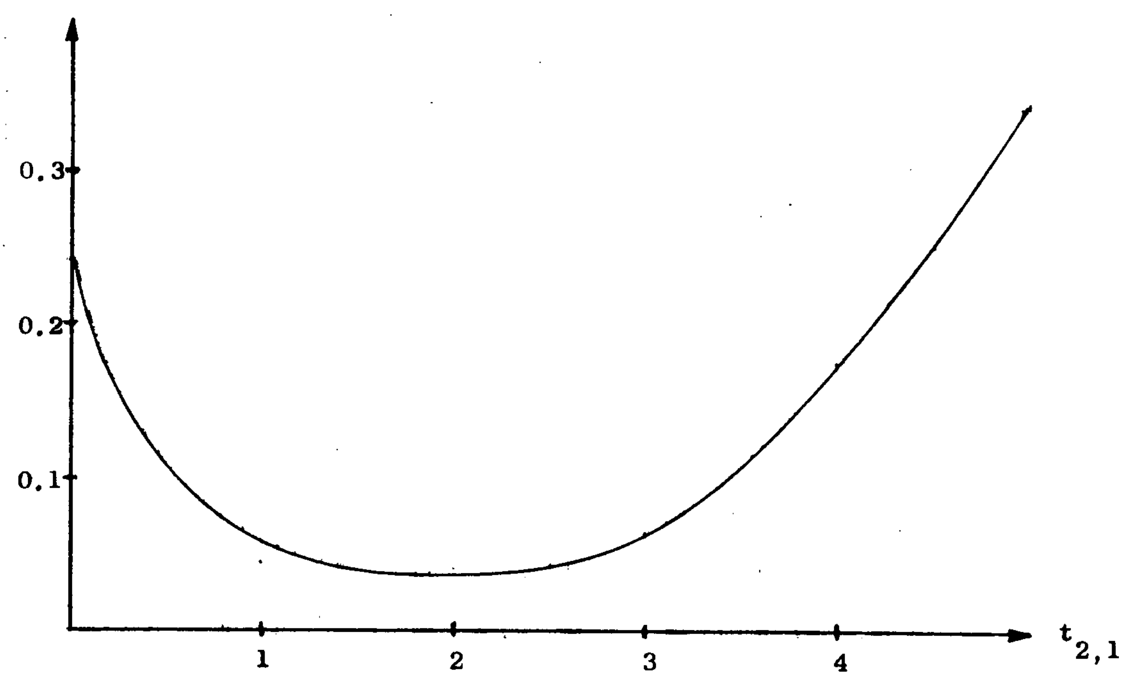


Fig. 9.2d: Case 4: $N_1 = 0$, $N_2 = 1$, $P = -1$

FIG. 9.2 (Cont) PERFORMANCE INDEX VS ONE SWITCH TIME

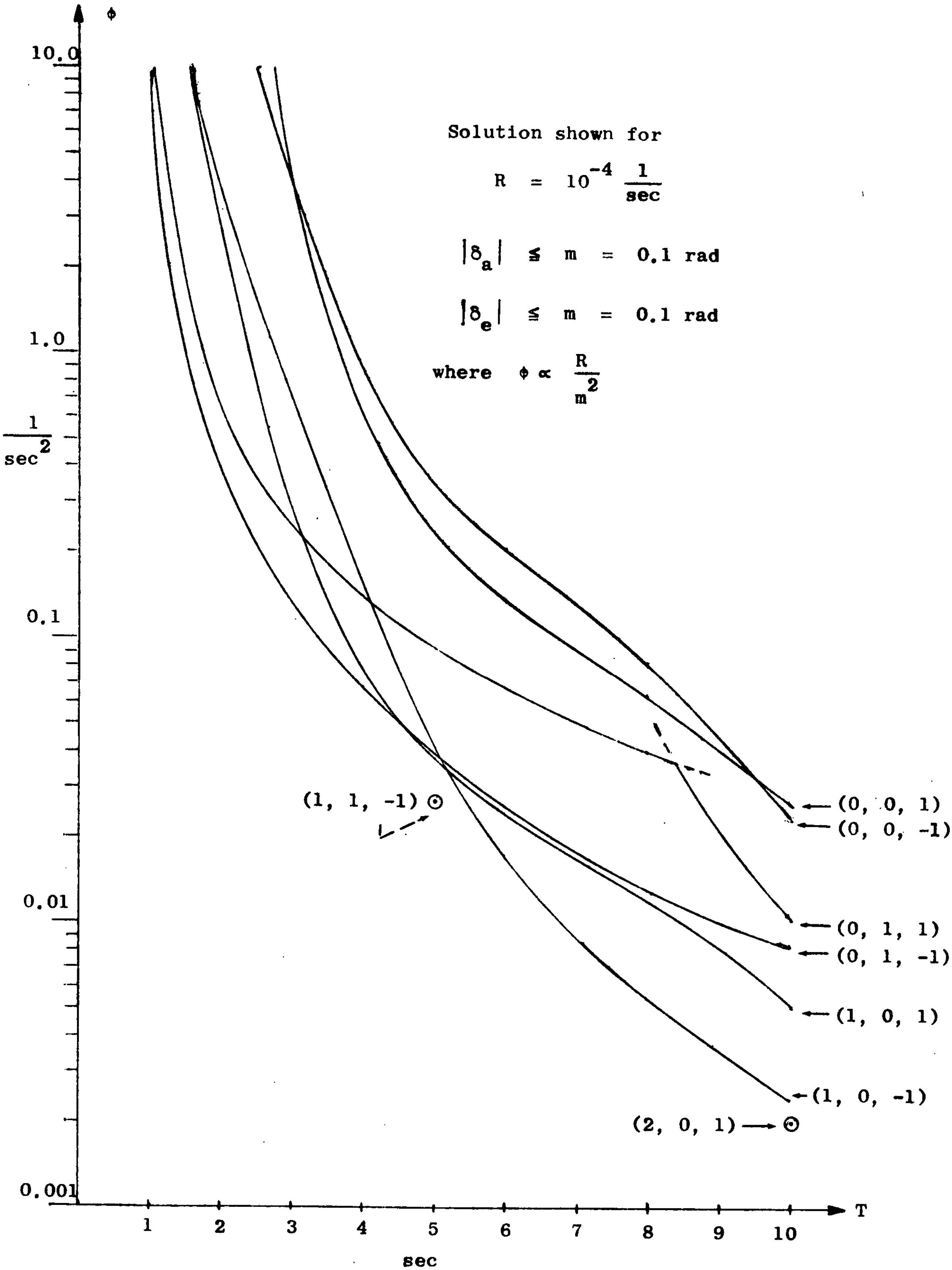


FIG. 9.3 PERFORMANCE INDEX VS LENGTH OF TEST

$t_{1,2}$ (sec) \ $t_{1,1}$ (sec)	0.0	1.0	2.0	3.0	4.0	5.0
5.0	0.346	0.061	0.036	0.048	0.135	0.236
4.0	0.368	0.070	0.043	0.064	0.236	
3.0	0.185	0.120	0.104	0.236		
2.0	0.049	0.410	0.236			
1.0	0.063	0.236				
0.0	0.236					

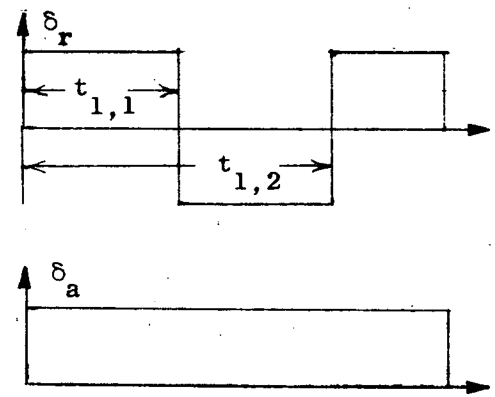


FIG. 9.4a PERFORMANCE INDEX VS ONE SWITCH TIME, CASE 1 (2, 0, +1)

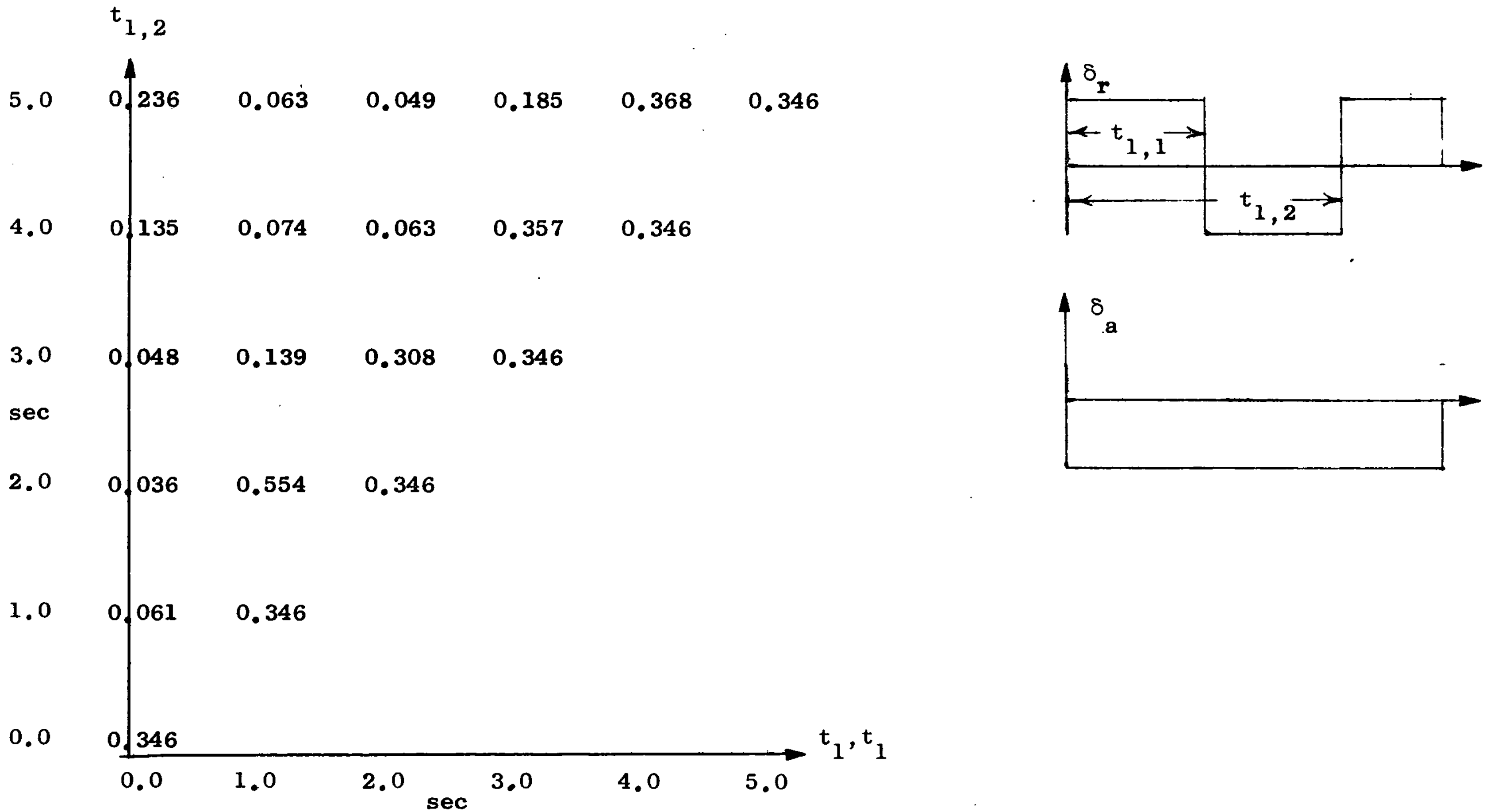


FIG. 9.4b PERFORMANCE INDEX VS TWO SWITCH TIMES, CASE 2 (2, 0, -1)

$t_{2,1}$ (sec) \ $t_{1,1}$ (sec)	0.0	1.0	2.0	3.0	4.0	5.0
5.0	0.346	0.061	0.036	0.048	0.135	0.236
4.0	0.179	0.059	0.039	0.059	0.213	0.391
3.0	0.066	0.058	0.064	0.163	1.912	4.329
2.0	0.042	0.086	0.275	0.349	0.313	0.310
1.0	0.044	1.549	0.133	0.071	0.083	0.093
0.0	0.236	0.063	0.049	0.185	0.368	0.346

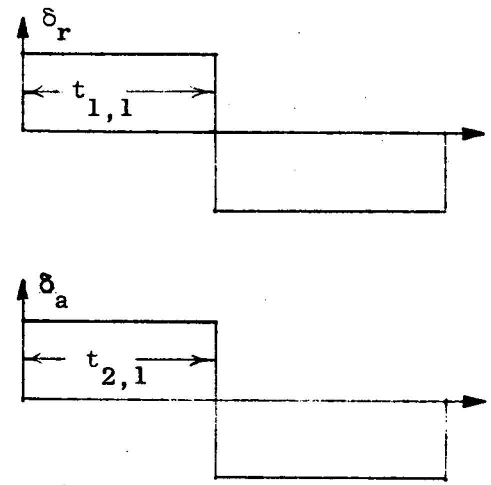


FIG. 9.4c PERFORMANCE INDEX VS TWO SWITCH TIMES, CASE 3 (1, 1, +1)

$t_{2,1}$ (sec) \ $t_{1,1}$ (sec)	0.0	1.0	2.0	3.0	4.0	5.0
5.0	0.236	0.063	0.049	0.185	0.368	0.346
4.0	0.391	0.061	0.069	0.536	0.281	0.179
3.0	4.329	0.061	0.405	0.207	0.074	0.066
2.0	0.310	0.121	0.193	0.042	0.040	0.042
1.0	0.093	0.312	0.034	⊙0.027 (global mimimum)	0.038	0.044
0.0	0.346	0.061	0.036	0.048	0.135	0.236

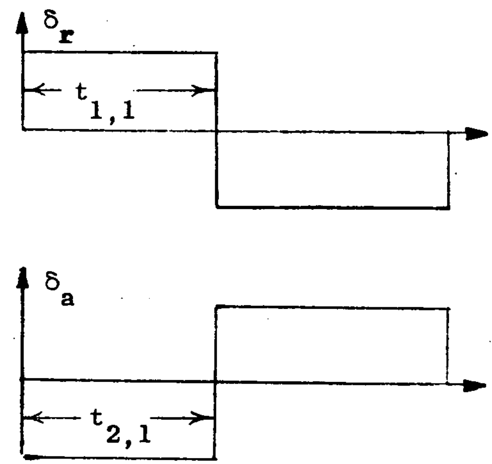


FIG. 9.4d PERFORMANCE INDEX VS TWO SWITCH TIMES, CASE 4 (1, 1, -1)

$t_{2,2}$ (sec) \ $t_{2,1}$ (sec)	0.0	1.0	2.0	3.0	4.0	5.0
5.0	0.346	0.093	0.310	4.329	0.391	0.236
4.0	0.179	0.082	0.313	2.158	0.236	
3.0	0.066	0.046	0.244	0.236		
2.0	0.042	0.056	0.236			
1.0	0.044	0.236				
0.0	0.236					

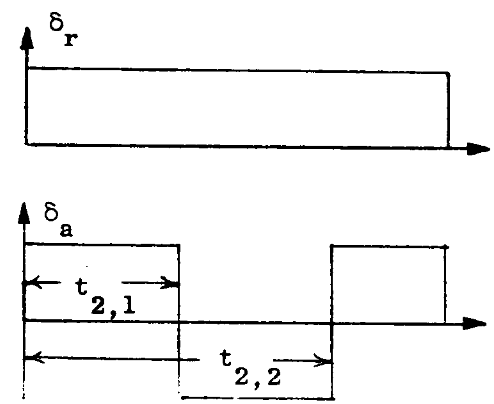
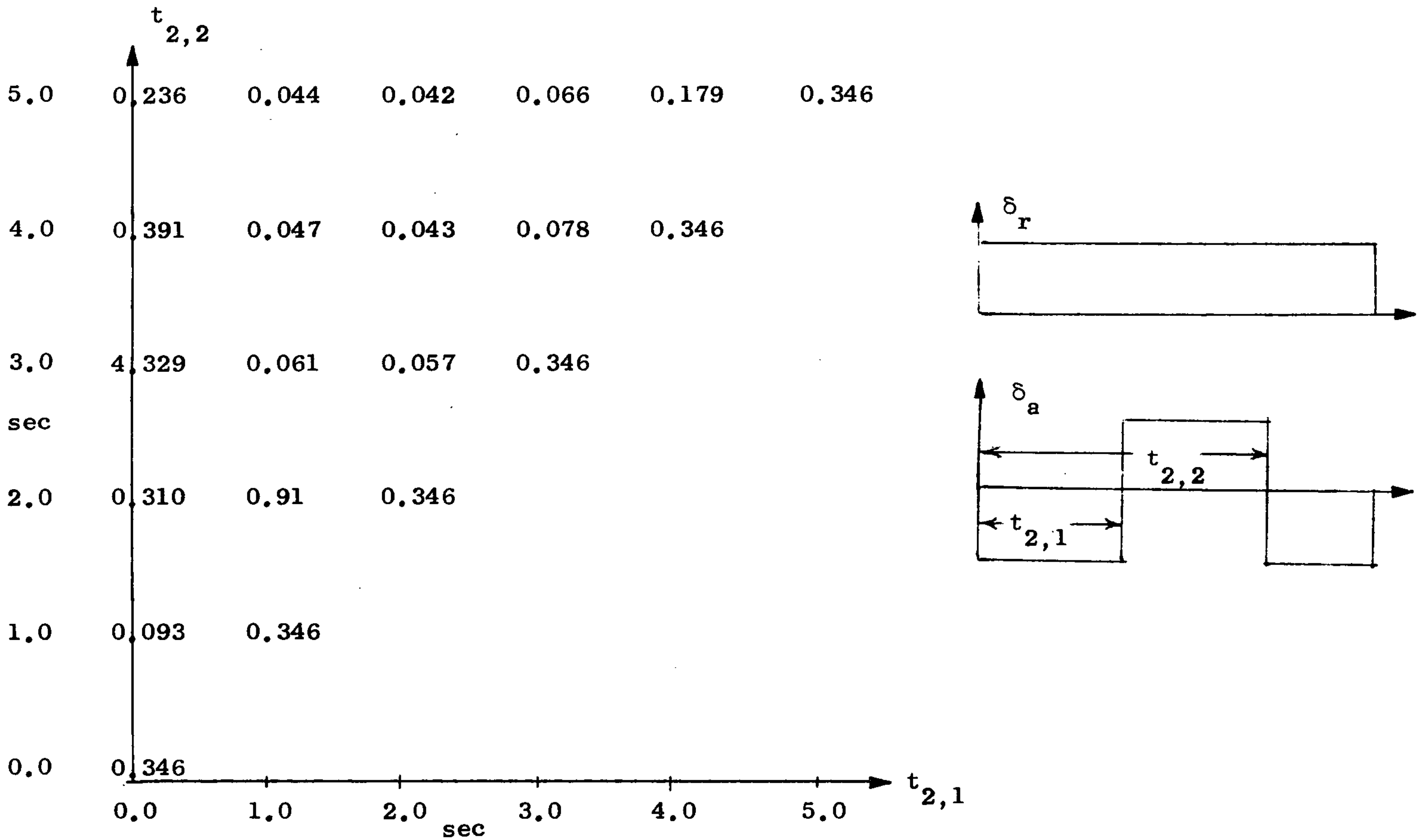


FIG. 9.4e PERFORMANCE INDEX VS TWO SWITCH TIMES, CASE 5 (0, 2, +1)



9.4f PERFORMANCE INDEX VS TWO SWITCH TIMES, CASE 6 (0, 2, -1)

All six cases for $N = 2$ were examined in a similar fashion for test lengths of one and ten sec. For the one sec test, each case had its minimum at an $N = 1$ case. For the ten sec test, each case had its minimum at an $N = 1$ case except for the (2, 0, 1) case. (The switch times for this case were 1.5 sec and 5.3 sec for the rudder, and no switching of the aileron.)

No $N = 3$ cases were investigated.

Unfortunately, these solutions cause such large deviations in the state that the linearity assumptions are violated. One method of satisfying the linearity requirement is the addition of state inequality constraints. For this problem this means two second-order state inequality constraints on β and ϕ .

With state inequality constraints, the steady state (assuming a stable system) wave shape may be somewhat irregular. A Fourier analysis may then be tried by optimizing with respect to the relative amplitude of higher order terms in addition to the frequency.

However, these problems are left for future research.

Chapter X

CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

A. CONCLUSIONS

Optimal input design for system identification has been investigated. The primary conclusions are:

1. The information matrix, I, (for the parameters of a linear dynamic system) provides a useful measure for input design. The criterion used in this thesis was the trace of I^{-1} (which is a lower bound of the covariance of the parameters). Minimizing this criterion appears to have some advantages over maximizing the trace of I. In simulations where the trace of I^{-1} was minimized, I^{-1} was a good lower bound in the sense that it was approximately equal to the actual covariance of the parameters.

2. An optimal input for system identification excites the system as much as possible. With amplitude constraints on the input, an optimal input is either full on in one direction, or full on in the opposite direction (bang-bang inputs). The addition of state inequality constraints can be important in practical problems where the instrumentation and the dynamics of the system must be maintained within their linear region. With the addition of state inequality constraints, the optimal input is still bang-bang but with intermediate values while on a state constraint.

3. For long tests, the optimal switch times are often equally spaced. In such cases, we may assume a square wave input and optimize the performance index with respect to the fundamental

frequency using a few terms of the Fourier series for a square wave. With state inequality constraints, the shape of the input pulses may require several terms in a Fourier series for an adequate approximation.

4. The results of a simple example indicate that for reasonable amounts of deterministic input, the overall effect of process noise is to decrease the identification accuracy. However, for systems with no (or very small) deterministic inputs, process noise contributes to the identification accuracy by providing excitation.

5. The solutions in this thesis for the optimal aircraft flight test may be modified to insure that the instrumentation and dynamics of the aircraft stay within their linear regions. One method of meeting the linearity requirement is to lower the input amplitude constraint. A design allowing full inputs but with switching to meet state-inequality constraints should prove better but has not been solved.

B. RECOMMENDATIONS

The following areas are recommendations for further research.

1. The methodology developed in this thesis should be extended to include the addition of state inequality constraints. Of immediate interest would be the addition of state inequality constraints to the aircraft identification problem.

2. The information matrix also provides a criterion for determining the best instrumentation to use. Instead of heavily instrumenting an aircraft or other system, it may be possible to obtain almost as much information with far less instrumentation. This would not only lower instrumentation costs but lower the complexity and execution time of identification algorithms. Identification algorithms could also be structured to process

only those measurements that contain the most information (at least for initial iterations). However, optimizing the best input/instrumentation combination together would be quite difficult.

3. As mentioned in Chapter III, more research would be useful in determining the best model numbers (numbers that specify structural information about the system such as order or degree of the minimal annihilation polynomial) for multi-input multi-output systems. Considerations should answer the following two questions: (a) What is the minimum number of parameters needed to designate an arbitrary member of the class defined by the model numbers? (b) As the order of the system increases, how many different cases must be examined?

4. The calculated value of the information matrix may vary with changes in the estimated value of the parameters. Instead of expanding the identification performance index to second order (as in Eq. 6.1), we could expand it to third or higher order. The third order tensor

$$E \frac{\partial^3 J}{\partial a^3}$$

may be viewed as the sensitivity of I with respect to the parameters. In addition to minimizing $\mathrm{Tr}\, I^{-1}$, some measure of this term should be minimized.

APPENDIX A

This Appendix is a computer listing of the gradient algorithm developed in Chapter VII, which is applied to the optimal input problem in Chapter VIII. A flow diagram of the conjugate gradient algorithm is shown in Fig. 7.1, and a flow diagram of the one dimensional search portion of the algorithm is shown in Fig. 7.2. Subroutine POINT calculates the value of the performance index by integrating the state, sensitivity, and elements of the information matrix, whose differential equations are in subroutine FCT. Subroutine GRAD calculates the partial derivatives of the performance index with respect to the switch times, which requires integrating the equations in subroutine FCTP. Subroutine ADAMS (not shown) was the numerical integration package used.

```
C     THE OPTIMUM INPUT TO IDENTIFY 2 PARAMETERS
C     (M ALPHA DOT AND M Q)  OF THE SHORT PERIOD DYNAMICS
C     OF AN AIRPLANE IS  A BANG-BANG INPUT
C     WITH SWITCH TIMES GIVEN BY THIS PROGRAM.
C     TO FIND THE GLOBAL MINIMUM OF JP, RUN THIS PROGRAM
C     SEVERAL TIMES WITH DIFFERENT VALUES FOR N AND PLOT
C     THE RESULTS.  ACCURACY IS CONTROLLED BY SPECIFING
C     IMAX,AMIN&SMIN
C     JP = TRACE OF THE COVARIANCE OF THE TWO PARAMETERS
C     N=NUMBER OF SWITCH TIME INTERVALS
C     T=L=LENGTH OF TEST
C     D(I)= I TH SWITCH TIME
C     IMAX=MAXIMUM NUMBER OF ITERATIONS
C     AMIN=MINIMUM INCREMENT TO ASSIGN TO SWITCH TIMES
C     SMIN=MINIMUM SUM OF SQUARES OF PJ'S
C     PJ=PARTIAL DER OF JP WITH RESPECT TO SWITCH TIMES
C     ZE=ZETA=DAMPING RATIO
C     WN=UNDAMPED NATERAL FREQUENCY
      REAL JP,JS,J0,J1,J2,J3
      EXTERNAL ADAMS
      DIMENSION D(10),PJ(10),X(15,11),H(10,10),R(10),
     C  DR(10),DG(10),DOLD(10),PJOLD(10),DX(10)
      COMMON C1,C2,U
      COMMON/S1/N,D,X,V1,V2,V3,DET
    1 FORMAT('1',35X,'ZETA=',F6.3,'  OMEGA N=',F6.3,
     C '  C1=',F6.3,'  C2=',F6.3)
    2 FORMAT(//' ',60X,'N=',I4/7X,'T',10X,'JP',9X,
     C 'D(2) THRU D(N+1)')
    3 FORMAT(' ',F9.3,F13.5,10F11.5)
  100 FORMAT(I10,2F10.5)
  101 FORMAT(2F10.5)
  102 FORMAT(7I3)
      READ(5,100) IMAX,AMIN,SMIN
      READ(5,101) ZE,WN
      READ(5,102) N1,N2,N3,L1,L2,L3
      C1=2.*ZE*WN-1.0
      C2=WN*WN-C1
      WRITE(6,1) ZE,WN,C1,C2
      DO 33 N=N1,N2,N3
      TN=N
      WRITE (6,2) N
      DO 32 L=L1,L2,L3
      T=L
      D(1)=0.0
      DO 5 I=2,N
    5 D(I)=D(I-1)+T/TN
      D(N+1)=T
      AT=0.2*(N-1)*T/TN
      ICNT=0
C
      CALL POINT(J0)
      IF (N.EQ.1) GO TO 30
      CALL GRAD(PJ)
```

```
    6 CONTINUE
      ICNT=ICNT+1
      K=1
      AOA=0.0
      DO 8 I=2,N
      DO 8 J=2,N
      H(I,J)=0.0
    8 IF(I.EQ.J) H(I,J)=1.0
      IF (ICNT.GE.IMAX) GO TO 30
    9 CONTINUE
      A1=AT
      DO 10 I=2,N
   10 DOLD(I)=D(I)
      RS=0.0
      PJ0=0.0
      DO 12 I=2,N
      R(I)=0.0
      DO 11 J=2,N
   11 R(I)=R(I)+H(I,J)*PJ(J)
      RS=RS+ABS(R(I))
   12 PJ0=PJ0+R(I)*PJ(I)
      PJ0=-PJ0/RS
C
C    ONE DIMENSIONAL SEARCH
C
      DO 13 I=2,N
   13 D(I)=D(I)-R(I)*A1/RS
      CALL POINT(J1)
      IF (J1.LE.J0+.83333333*PJ0*A1) GO TO 15
      A2=-.5*PJ0*A1*A1/(J1-J0-PJ0*A1)
      DEC=-.5*PJ0*A2
      DO 14 I=2,N
   14 D(I)=D(I)-R(I)*(A2-A1)/RS
      CALL POINT(J2)
      IF(ABS((J0-J2-DEC)/DEC).GE.0.1) GO TO 17
      GO TO 20
   15 A2=3.0*A1
      DO 16 I=2,N
   16 D(I)=D(I)-R(I)*(A2-A1)/RS
      CALL POINT(J2)
   17 E1=(J1-J0-PJ0*A1)/(A1*A1)
      E2=(J2-J0-PJ0*A2)/(A2*A2)
      E4=(E1-E2)/(A1-A2)
      E3=E1-E4*A1
      IF (E3*E3-3.0*PJ0*E4 .LT. 0.0) GO TO 20
      A3=(-E3+SQRT(E3*E3-3.0*PJ0*E4))/(3.0*E4)
      IF (A3.LE.0.0) GO TO 20
      IF (A3.GE.6.0*A1) A3=6.0*A1
      DO 18 I=2,N
   18 D(I)=D(I)-R(I)*(A3    -A2)/RS
      CALL POINT (J3)
      IF (J3.LE.J2) GO TO 22
```

```
      DO 19 I=2,N
   19 D(I)=D(I)-R(I)*(A2-A3)/RS
      CALL POINT(J2)
   20 AOP=A2
      JO=J2
      GO TO 23
   22 CONTINUE
      AOP=A3
      JO=J3
   23 CONTINUE
C
C     END OF ONE DIMENSIONAL SEARCH
C
      IF (K.EQ.1 .AND. AOP.LE.AMIN) GO TO 30
      DO 24 I=2,N
   24 PJOLD(I)=PJ(I)
      CALL GRAD(PJ)
      S=0.0
      DO 25 I=2,N
   25 S=S+PJ(I)*PJ(I)
      IF(S.LE.SMIN) GO TO 30
      IF(AOP.GE.3.0*A1.AND.ICNT.EQ.1.AND.K.EQ.1) GO TO 9
      K=K+1
      AOA=AOA+0.2*AOP/(N-1)
      IF (K.GE.N) AT=AOA
      IF (K.GE.N) GO TO 6
C
C     CALCULATE H MATRIX
      DO 26 I=2,N
      DX(I)=D(I)-DOLD(I)
   26 DG(I)=PJ(I)-PJOLD(I)
      DO 27 I=2,N
      DR(I)=0.0
      DO 27 J=2,N
   27 DR(I)=DR(I)+H(I,J)*DG(J)
      DM1=0.0
      DM2=0.0
      DO 28 I=2,N
      DM1=DM1+DX(I)*DG(I)
   28 DM2=DM2+DG(I)*DR(I)
      DO 29 I=2,N
      DO 29 J=2,N
   29 H(I,J)=H(I,J)+DX(I)*DX(J)/DM1-DR(I)*DR(J)/DM2
C
      GO TO 9
   30 CONTINUE
      IN=N+1
      WRITE (6,3) T,JO,(D(I),I=2,IN)
   32 CONTINUE
   33 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE POINT(JP)
C     VALUES OF X AT THE SWITCH TIMES AND VALUE OF JP
      EXTERNAL FCT
      REAL JP
      DIMENSION D(10),X(15,11),XI(15),XF(15)
      COMMON C1,C2,U
      COMMON/S1/N,D,X,V1,V2,V3,DET
      NN=9
      DO 1 K=1,NN
      X(K,1)=0.0
    1 XI(K)=0.0
      DO  3 I=1,N
      U=(-1)**(I+1)
      CALL ADAMS(NN+1,D(I),D(I+1),XI,XF,FCT)
      DO  2 J=1,NN
      XI(J)=XF(J)
    2 X(J,I+1)=XF(J)
    3 CONTINUE
      V1=X(7,N+1)
      V2=X(8,N+1)
      V3=X(9,N+1)
      DET=V1*V2-V3*V3
      P11=V1/DET
      P12=-V3/DET
      P22=V2/DET
      JP=P11+P22
      RETURN
      END
```

```
      SUBROUTINE GRAD(PJ)
C PARTIAL DERIVATIVES OF X WITH RESPECT TO SWITCH TIMES
      EXTERNAL FCTP
      DIMENSION D(10),X(15,11),PJ(10),XI(15),XF(15)
      COMMON C1,C2,U
      COMMON/S1/N,D,X,V1,V2,V3,DET
      NN=15
      DO 5  J=2,N
      U=(-1)**(J+1)
      DO  1 K=1,6
    1 XI(K)=X(K,J)
      XI(7)=-2.*U
      DO  2 K=8,15
    2 XI(K)=0.0
      II=N-J +1
      DO  4 I=1,II
      CALL ADAMS(NN+1,D(I+J-1),D(I+J),XI,XF,FCTP)
      DO  3 K=1,NN
    3 XI(K)=XF(K)
      U=-U
    4 CONTINUE
      PV1=XI(13)
      PV2=XI(14)
      PV3=XI(15)
      PJ(J)=(-PV1*(V2*V2+V3*V3)-PV2*(V1*V1+V3*V3)
     C +2.*PV3*V3*(V1+V2))/DET**2
    5 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE FCT(T,X,DX)
C     DIFFERENTIAL EQUATIONS FOR STATE,SENSITIVITY
C     EQUATIONS AND INFORMATION MATRIX
      DIMENSION X(15),DX(15)
      COMMON C1,C2,U
      DX(1)=-C1*X(1)-C2*X(2)+U
      DX(2)=X(1)-X(2)
      DX(3)=-C1*X(3)-C2*X(4)+X(1)-X(2)
      DX(4)=X(3)-X(4)
      DX(5)=-C1*X(5)-C2*X(6)+X(1)
      DX(6)=X(5)-X(6)
      DX(7)=X(5)*X(5)
      DX(8)=X(3)*X(3)
      DX(9)=X(3)*X(5)
      RETURN
      END
```

```
      SUBROUTINE FCTP(T,X,DX)
C     DIFFERENTIAL EQUATIONS FOR STATE,SENSITIVITY
C     EQUATIONS, THEIR DERIVATIVES WITH RESPECT TO
C     SWITCH TIMES AND DERIVATIVES OF INFORMATION MATRIX
      DIMENSION X(15),DX(15)
      COMMON C1,C2,U
      DX(1)=-C1*X(1)-C2*X(2)+U
      DX(2)=X(1)-X(2)
      DX(3)=-C1*X(3)-C2*X(4)+X(1)-X(2)
      DX(4)=X(3)-X(4)
      DX(5)=-C1*X(5)-C2*X(6)+X(1)
      DX(6)=X(5)-X(6)
      DX(7)=-C1*X(7)-C2*X(8)
      DX(8)=X(7)-X(8)
      DX(9)=-C1*X(9)-C2*X(10)+X(7)-X(8)
      DX(10)=X(9)-X(10)
      DX(11)=-C1*X(11)-C2*X(12)+X(7)
      DX(12)=X(11)-X(12)
      DX(13)=2.*X(5)*X(11)
      DX(14)=2.*X(3)*X(9)
      DX(15)=    X(3)*X(11)+X(5)*X(9)
      RETURN
      END
```

APPENDIX B

This Appendix consists of two parts. The first part is a computer listing of the simulation algorithm developed in Chapter VIII. The simulation consists of applying Denery's combined algorithm to repeated sets of simulated data and calculating the covariance of the resulting estimates. The second part is a listing from the simulation program for the optimal input case for a set of 20 tests. Of special note are the last three columns which (when multiplied by $R = 10^{-4}$) show values of I_a^{-1} based upon the estimated values of the parameters. These values ranged from slightly under the true covariance (shown in the last line) to 50% over the true covariance, and indicate the sensitivity of the information matrix with respect to errors in the estimates of the parameters.

```
C     IDENTIFICATION SIMULATICN USING THE OPTIMUM INPUT
C     TO IDENTIFY M/ADOT AND M/Q FROM THE SHORT PERIOD
C     DYNAMICS OF AN AIRPLANE
      REAL K32,K34,KN32,KN34,K54,M1,K11,K21,KR11,
     C KR21,K14
      DIMENSICN U(1000),Y(1000),XI( 50),XF( 50),
     C Z(1000),SW(10),ST(6,100),SUM(10)
      COMMON U,Y,SWT,F11,F21,FN11,FN21,G11,G21,Z,
     C K14,K54,G,K11,K34,SW,NN
      EXTERNAL FCT1,FCT2
    1 FORMAT(8F10.5)
    2 FORMAT(3I3,1X,I10,2F10.5)
    3 FORMAT('1',30X,' IDENTIFICATION ALGORITHM',
     C ' WITH OPTIMAL INPUTS'/' N=',I3,' NT=',I3,
     C ' IP=',I3,' IX=',I10,' S=',F10.5,' ACC=',F10.5)
    4 FORMAT(' SWITCH TIMES ARE ',10F11.5)
    5 FORMAT(' K54=',F10.5,' G11=',F10.5,' G21=',
     C F10.5,' M1=',F10.5,'K11=',F10.5,' K14=',F10.5
     C /' SIMULATED (ACTUAL) VALUES OF THE UNKNOWN ',
     C ' CONSTANTS ARE ON THE FIRST LINE'/' NOMINAL ',
     C 'STARTING VALUES ARE ON THE SECOND LINE'/
     C ' VALUES USING DENERY"S ALGORITHM ARE ON SUB
     CSEQUENT LINES')
    6 FORMAT(//40X,' TEST NUMBER',I3/
     2'PK22',7X,'P1',8X,'P2',8X,'P11',8X,'P12',8X,'P22')
    7 FORMAT(' ',I4,2(2F10.6,3F11.6))
    8 FORMAT(//' K-STATISTICS CALCULATED FROM THE ABCVE '
     1,I3,' TESTS'/9X,'F11',7X,'F21',7X,'PK21',7X,'PK22',
     27X,'P1',8X,'P2',8X,'P11',8X,'P12',8X,'P22')
    9 FORMAT(5X,2(2F10.6,3F11.6))
   17 FORMAT(' ',I4,2F10.6,33X,2F10.6)
C     INITIALIZATION
      AM=0.0
      DO 30 I=1,10
   30 SUM(I)=0.0
      DO 10 I=1,50
   10 XI(I)=0.0
C     READ IN FOLLOWING PARAMETERS
C     N=  NUMBER OF SWITCH TIMES +1
C     NT= NUMBER OF TESTS
C     IP= PRINT OPTION
C     IX= RANDOM NUMBER
C     S=  STANDARD DEV OF MEASUREMENTS
C     ACC=REQUIRED ACCURACY OF ID ALGORITHM
C     T=  LENGTH OF TEST
C     SW= SWITCH TIMES
      READ (5,2) N,NT,IP,IX,S,ACC
      WRITE(6,3) N,NT,IP,IX,S,ACC
      READ(5,1) (SW(I),I=1,N)
      WRITE(6,4) (SW(I),I=1,N)
      T=SW(N)
      NN=N-1
C     KNOWN CONSTANTS FOR THE C-8 AIRPLANE
      G=32.16
      READ (5,1) K54,G11,M1,K11,K14
      G21=-K54*G11
```

```
      WRITE(6,5) K54,G11,G21,M1,K11,K14
C     SIMULATED VALUES FOR UNKNOWN CONSTANTS
      READ (5,1) P1,P2
      K32=P1+P2
      K34=P1*K54 +M1
      F11=K32+K54
      F21=K34-K54*K32
C     NOMINAL VALUES FOR UNKNOWN CONSTANTS
      READ (5,1) PS1,PS2
C     A CALL TO ADAMS WITH SUBROUTINE FCT1
C     GENERATES TRUE INPUT AND OUTPUT
      CALL ADAMS(5,0.0,T   ,XI,XF,FCT1,IP,1)
C     ALGORITHM REPEATED ON NT SETS OF DATA
      DO 21 K=1,NT
      ICNT=0
      WRITE(6,6) K
      WRITE(6,17) ICNT,F11,F21,P1,P2
      PN1=PS1
      PN2=PS2
      KN32=PN1+PN2
      KN34=PN1*K54+M1
      FN11=KN32+K54
      FN21=KN34-K54*KN32
      WRITE(6,17) ICNT,FN11,FN21,PN1,PN2
C     NORMAL RANDOM NUMBER ADDED TO MEASUREMENT
      DO 11 I=1,401
      A=0.0
      DO 50 J=1,12
      IY=IX*65539
      IF(IY) 55,56,56
   55 IY=IY+2147483647+1
   56 YFL=IY
      YFL=YFL*.4656613E-9
      IX=IY
   50 A=A+YFL
      V=(A-6.0)*S+AM
   11 Z(I)=Y(I)+V
C     IDENTIFICATION ALGORITHM
      SWT=1.0
   15 ICNT=ICNT+1
      IF (ICNT.GE.10) GO TO 20
C     A CALL TO ADAMS WITH FCT2 GENERATES NOMINAL OUTPUT,
C     SENSITIVITY EQUATIONS AND NECESSARY QUADRATURES
      CALL ADAMS(12,0.0,T   ,XI,XF,FCT2,0,1)
      W1=XF(7)
      W2=XF(8)
      V1=XF(9)
      V2=XF(10)
      V3=XF(11)
      DET=V1*V2-V3*V3
      PK11=V1/DET
      PK12=-V3/DET
      PK22=V2/DET
      KR11=(V1*W1-V3*W2)/DET
      KR21=(-V3*W1+V2*W2)/DET
      FN11=FN11+KR11
```

```
      FN21=FN21+KR21
      KN32=FN11-K54
      KN34=FN21+K54*KN32
      PN1=(KN34-M1)/K54
      PN2=KN32-PN1
      P11=PK11+2.*PK12/K54+PK22/K54**2
      P12=-PK12/K54-PK22/K54**2
      P22=PK22/K54**2
      WRITE(6,7) ICNT,FN11,FN21,PK11,PK12,PK22,PN1,PN2,
     CP11,P12,P22
C     IF CHANGES IN ESTIMATES ARE LESS THAN ACC THEN PROCEED
C     TO STEP 2 OR IF ON STEP 2 STOP
      IF (ABS(KR11).LT.ACC .AND.ABS(KR21).LT.ACC) GO TO 16
      GO TO 15
   16 IF (SWT.LT.0.0) GO TO 20
      ICNT=0
      SWT=-1.0
      GO TO 15
   20 CONTINUE
C     STORE ESTIMATES FOR LATER ANALYSIS
      ST(1,K)=FN11
      ST(2,K)=FN21
      ST(3,K)=PN1
      ST(4,K)=PN2
      SUM(1)=SUM(1)+FN11
      SUM(2)=SUM(2)+FN21
      SUM(6)=SUM(6)+PN1
      SUM(7)=SUM(7)+PN2
   21 CONTINUE
C     CALCULATE THE ACTUAL MEAN AND COVARIANCE
      SUM(1)=SUM(1)/NT
      SUM(2)=SUM(2)/NT
      SUM(6)=SUM(6)/NT
      SUM(7)=SUM(7)/NT
      IF(NT.EQ.1) GO TO 23
      WRITE(6,8) NT
      DO 22 J=1,NT
      SUM(3)=SUM(3)+(SUM(1)-ST(1,J))**2
      SUM(4)=SUM(4)+(SUM(1)-ST(1,J))*(SUM(2)-ST(2,J))
      SUM(5)=SUM(5)+(SUM(2)-ST(2,J))**2
      SUM(8)=SUM(8)+(SUM(6)-ST(3,J))**2
      SUM(9)=SUM(9)+(SUM(6)-ST(3,J))*(SUM(7)-ST(4,J))
      SUM(10)=SUM(10)+(SUM(7)-ST(4,J))**2
   22 CONTINUE
      SUM(3 )=SUM(3 )/(NT-1.)
      SUM(4 )=SUM(4 )/(NT-1.)
      SUM(5 )=SUM(5 )/(NT-1.)
      SUM(8 )=SUM(8 )/(NT-1.)
      SUM(9 )=SUM(9 )/(NT-1.)
      SUM(10)=SUM(10)/(NT-1.)
      WRITE(6,9) (SUM(I),I=1,10)
   23 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE FCT1(T,X,DX)
C     GENERATES SIMULATED INPUT AND OUTPUT MEASUREMENTS
      REAL K32,K34,KN32,KN34,K54,M1,K11,K21,KR11,KR21,K14
      DIMENSION X( 50),DX( 50),U(1000),Y(1000),Z(1000),SW(10)
      COMMON U,Y,SWT,F11,F21,FN11,FN21,G11,G21,Z,
     C K14,K54,G,K11,K34,SW,NN
      I=INT(100.01*T) +1
      U(I)=0.2
      DO 10 J=1,NN,2
   10 IF(T.GE.SW(J).AND.T.LT.SW(J+1)) U(I)=-0.2
      DX(1)=F11*X(1)+X(2)+G11*U(I)
      DX(2)=F21*X(1)      +G21*U(I)
      DX(3)=X(1)
      DX(4)=K11*X(4)-G*X(3)+K14*(K54*X(1)+X(2))/K34
      Y(I)=X(1)
      RETURN
      END
```

```
      SUBROUTINE FCT2(T,X,DX)
C     GENERATES NOMINAL OUTPUT AND SENSITIVITY EQUATIONS.
C     PERFORMS QUADRATURES.
      REAL K32,K34,KN32,KN34,K54,M1,K11,K21,KR11,KR21,K14
      DIMENSION X( 50),DX( 50),U(1000),Y(1000),Z(1000),SW(10)
      COMMON U,Y,SWT,F11,F21,FN11,FN21,G11,G21,Z,
     C K14,K54,G,K11,K34,SW,NN
      I=INT(100.01*T) +1
      DX(1)=FN11*X(1)+X(2)+G11*U(I)
      DX(2)=FN21*X(1)      +G21*U(I)
      YD=X(1)
      IF (SWT.GT.0.0) YD=Z(I)
      DX(3)=FN11*X(3)+X(4)+YD
      DX(4)=FN21*X(3)
      DX(5)=FN11*X(5)+X(6)
      DX(6)=FN21*X(5)+YD
      DX(7)=X(3)*(Z(I)-X(1))
      DX(8)=X(5)*(Z(I)-X(1))
      DX(9)=X(5)*X(5)
      DX(10)=X(3)*X(3)
      DX(11)= X(3)*X(5)
      RETURN
      END
```

IDENTIFICATION ALGORITHM WITH OPTIMAL INPUTS

N= 2 NT= 20 IP= 0 IX= 8642571 S= 0.10000 ACC= 0.00010
SWITCH TIMES ARE 2.72000 4.00000
K54= -0.73700 G11= -1.65800 G21= -1.22195 M1= -1.14718 K11= -0.02000 K14= 33.73599
SIMULATED (ACTUAL) VALUES OF THE UNKNOWN CONSTANTS ARE ON THE FIRST LINE
NOMINAL STARTING VALUES ARE ON THE SECOND LINE
VALUES USING DENERY"S ALGORITHM ARE ON SUBSEQUENT LINES

TEST NUMBER 1

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.575132	-1.765865	143.308960	-2.678180	86.636978	-0.998664	-0.839468	310.079346	-153.136688	159.502808
2	-2.580450	-1.758019	187.007721	-4.608769	105.542740	-1.014627	-0.828823	393.823730	-200.562637	194.309219
3	-2.580305	-1.757697	187.413330	-4.718835	105.256958	-1.014918	-0.823386	394.001709	-200.185837	193.783081
4	-2.580293	-1.757690	187.391678	-4.721217	105.230209	-1.014916	-0.828376	393.937500	-200.139832	193.733841
1	-2.584389	-1.755280	191.162231	-5.020158	110.734012	-1.022283	-0.825106	408.743896	-210.770248	203.958649
2	-2.584394	-1.755255	191.956039	-5.121044	110.820129	-1.022322	-0.825071	409.878174	-210.973633	204.025146

TEST NUMBER 2

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.371102	-1.826491	129.856369	-1.825404	87.611633	-0.712373	-0.921729	296.106934	-163.773987	161.297195
2	-2.370629	-1.826750	151.385330	-2.002089	102.409134	-0.711550	-0.922079	345.358398	-191.256638	188.540100
3	-2.370625	-1.826752	151.346756	-1.998884	102.405640	-0.711542	-0.922082	345.304688	-191.245850	188.533661
1	-2.378772	-1.821386	152.064804	-1.928270	102.024918	-0.726969	-0.914802	345.130127	-190.449097	187.832733
2	-2.378819	-1.821304	153.371796	-2.053610	102.014999	-0.727128	-0.914691	346.759033	-190.600922	187.814484

TEST NUMBER 3

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.475343	-1.687908	133.456299	-2.951298	81.954559	-1.004649	-0.733693	292.347412	-154.886719	150.882248
2	-2.478008	-1.685228	162.816788	-5.087181	91.530783	-1.010951	-0.730057	345.323730	-175.601898	168.696640
3	-2.478018	-1.685178	163.024536	-5.131815	91.579351	-1.011030	-0.729987	345.552734	-175.565079	168.601974
1	-2.468043	-1.683057	167.729202	-5.335139	93.436859	-1.003933	-0.727110	354.228760	-179.260696	172.021713
2	-2.468243	-1.683061	165.730011	-5.260556	92.429993	-1.004127	-0.727115	350.173584	-177.305817	170.168030
3	-2.468239	-1.683062	165.768539	-5.262742	92.445724	-1.004122	-0.727117	350.245326	-177.337753	170.196991

TEST NUMBER 8

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.284647	-1.856217	120.559219	-1.280013	86.242981	-0.585584	-0.952062	282.810059	-160.514221	156.777435
2	-2.286882	-1.857075	133.709457	-1.017851	99.123103	-0.586656	-0.963226	318.961914	-183.871445	182.490372
3	-2.286925	-1.857093	133.901566	-1.018634	99.266541	-0.586674	-0.963251	319.420166	-184.136566	182.754440
1	-2.284150	-1.852517	138.385620	-1.114999	99.025864	-0.590108	-0.957041	323.722656	-183.824219	182.311340
2	-2.284099	-1.852555	137.780502	-1.147380	98.236359	-0.590005	-0.957093	321.751953	-182.414642	180.857819

TEST NUMBER 9

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.433477	-1.692291	125.637634	-2.613599	85.290253	-0.956838	-0.739639	289.753418	-150.569687	157.023422
2	-2.444534	-1.697599	150.590363	-4.323940	94.545303	-0.960693	-0.746841	336.386475	-179.929352	174.062408
3	-2.444704	-1.697713	151.648666	-4.331157	95.294540	-0.960708	-0.746995	338.843750	-181.318512	175.441788
4	-2.444708	-1.697715	151.665543	-4.330857	95.307724	-0.960710	-0.746998	338.884033	-181.342392	175.466064
1	-2.471483	-1.692807	161.727478	-4.695725	92.326385	-0.994144	-0.740339	344.447510	-176.348679	169.977280
2	-2.471676	-1.692703	166.702698	-5.098780	93.814117	-0.994478	-0.740198	353.255371	-179.634552	172.716263
3	-2.471676	-1.692704	166.736938	-5.103216	93.817245	-0.994477	-0.740199	353.307373	-179.646317	172.722015

TEST NUMBER 10

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.289425	-1.863508	117.775528	-1.161313	82.820740	-0.579470	-0.971955	273.403809	-154.052658	152.476929
2	-2.293766	-1.865191	131.381104	-0.905633	95.801437	-0.582528	-0.974238	310.213623	-177.603821	176.375015
3	-2.293921	-1.865171	131.838181	-0.909387	96.111191	-0.582709	-0.974212	311.251221	-178.179184	176.945282
4	-2.293926	-1.865169	131.850693	-0.910006	96.116394	-0.582718	-0.974208	311.274902	-178.189606	176.954865
1	-2.297057	-1.861901	139.816833	-1.068251	100.603439	-0.590282	-0.959775	327.931396	-186.665176	185.215729
2	-2.297050	-1.861866	140.248489	-1.114726	100.446915	-0.590322	-0.959727	328.200928	-186.440079	184.927567

TEST NUMBER 11

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.188787	-1.905384	114.597244	-0.892565	81.917725	-0.423011	-1.028775	267.833740	-152.025513	150.814438
2	-2.191311	-1.909681	120.756470	-0.376574	93.433502	-0.419705	-1.034605	293.793701	-172.526474	172.015533
3	-2.191385	-1.909767	121.011658	-0.367716	93.735343	-0.419663	-1.034721	294.673584	-173.163147	172.664215
1	-2.190832	-1.905387	124.478424	-0.368123	97.881546	-0.425053	-1.028778	305.631885	-180.704071	180.204590
2	-2.190894	-1.905321	124.265320	-0.392182	97.296509	-0.425204	-1.028689	304.457031	-179.659637	179.127518

TEST NUMBER 4

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.325520	-1.701275	122.978165	-2.303796	79.657913	-0.836691	-0.751828	275.883789	-149.779922	146.654022
2	-2.328340	-1.702418	137.696655	-3.293389	84.457947	-0.837959	-0.753380	302.125000	-159.959732	155.491104
3	-2.328370	-1.702429	137.944901	-3.295957	84.620316	-0.837976	-0.753394	302.673955	-160.262161	155.790039
1	-2.325012	-1.700657	140.942001	-3.451774	84.258035	-0.837022	-0.750990	305.450439	-159.824997	155.141464
2	-2.324994	-1.700653	140.321671	-3.451281	83.839584	-0.836996	-0.750998	304.040039	-159.035553	154.352575

TEST NUMBER 5

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.233837	-1.652736	119.383148	-2.483619	74.036227	-0.810869	-0.685968	262.519043	-139.766159	136.396271
2	-2.233345	-1.652827	125.676514	-3.488634	73.103378	-0.810252	-0.686093	269.730225	-139.320358	134.586807
3	-2.233349	-1.652828	125.639725	-3.485826	73.091873	-0.810254	-0.686094	269.664795	-139.295364	134.565628
1	-2.233424	-1.650047	124.212830	-3.423422	72.817551	-0.814104	-0.682320	267.563477	-138.705643	134.060577
2	-2.233425	-1.650048	124.157700	-3.468212	72.557602	-0.814103	-0.682322	267.151367	-138.287842	133.582001

TEST NUMBER 6

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736775				-0.600000	-0.800000			
1	-2.355214	-1.693687	133.718994	-3.000283	85.315757	-0.876680	-0.741533	298.932861	-161.143158	157.072220
2	-2.353938	-1.701896	152.067657	-4.386977	91.171509	-0.864267	-0.752671	331.823486	-173.303558	167.851089
3	-2.353744	-1.701526	152.064606	-4.302353	91.592331	-0.864574	-0.752169	332.365723	-174.463501	168.625854
4	-2.353756	-1.701535	152.042694	-4.305185	91.563354	-0.864574	-0.752182	332.298096	-174.413986	168.572495
1	-2.353107	-1.699333	145.297653	-3.698842	85.995880	-0.866913	-0.749194	313.657715	-163.341278	158.322510
2	-2.353059	-1.699374	145.122070	-3.730597	85.709375	-0.866908	-0.749250	313.041748	-162.858017	157.796158

TEST NUMBER 7

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736775				-0.600000	-0.800000			
1	-2.213266	-1.773271	126.440781	-2.113433	83.306549	-0.626750	-0.849516	285.547119	-156.238937	153.371323
2	-2.209773	-1.773498	133.013931	-2.122506	88.165298	-0.622949	-0.849824	301.090088	-165.195426	162.316513
3	-2.209902	-1.773440	132.735962	-2.109680	88.047397	-0.623154	-0.849747	300.561279	-164.962891	162.100372
4	-2.209897	-1.773444	132.745590	-2.110481	88.049683	-0.623145	-0.849751	300.576416	-164.967270	162.103668
1	-2.220175	-1.769747	123.594513	-1.609366	83.485107	-0.638439	-0.844735	281.661865	-155.883728	153.700058
2	-2.220163	-1.769770	125.093826	-1.710092	83.816162	-0.638398	-0.844765	284.043945	-156.629883	154.309555

TEST NUMBER 12

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.232770	-1.548800	106.704391	-2.604049	66.780223	-0.950326	-0.544943	236.716583	-126.478882	122.945587
2	-2.238173	-1.543928	111.681473	-4.435946	61.783356	-0.962841	-0.538332	237.465408	-119.765015	113.746094
3	-2.238437	-1.543785	112.025391	-4.504532	61.744293	-0.963298	-0.538139	237.923523	-119.786148	113.674179
4	-2.238449	-1.543781	112.042953	-4.507106	61.746170	-0.963316	-0.538133	237.951538	-119.793106	113.677643
1	-2.262530	-1.537107	122.484665	-5.387023	63.413864	-0.996453	-0.529077	253.851379	-124.057327	116.747940
2	-2.262831	-1.536901	126.052063	-5.795784	64.261948	-0.997032	-0.528798	250.039355	-125.173325	118.309311
3	-2.262834	-1.536900	125.094452	-5.803203	64.262451	-0.997037	-0.528796	260.152832	-126.184311	118.310226

TEST NUMBER 13

ITEP	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.377703	-1.816300	126.670517	-1.749697	90.536380	-0.732801	-0.907901	298.101318	-169.056778	166.682709
2	-2.381046	-1.819800	147.755127	-1.961792	105.487350	-0.731396	-0.912650	347.285889	-196.869095	194.207245
3	-2.381035	-1.819797	148.092300	-1.949160	105.856476	-0.731389	-0.912645	348.269555	-197.531540	194.886825
1	-2.367290	-1.818245	153.733932	-2.090427	101.999451	-0.719750	-0.910540	347.192383	-190.622253	187.785858
2	-2.367343	-1.818239	151.192108	-2.018483	100.681213	-0.719811	-0.910532	342.028564	-189.097687	185.358917

TEST NUMBER 14

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.267365	-1.690999	129.175385	-2.772858	75.335594	-0.792479	-0.737885	275.398193	-142.460602	138.698257
2	-2.262108	-1.688504	139.312759	-3.717070	77.250595	-0.790606	-0.734501	291.521582	-147.265533	142.222031
3	-2.262210	-1.688541	139.826080	-3.712728	76.951447	-0.790658	-0.734551	290.572510	-146.708893	141.671290
4	-2.262209	-1.688542	139.834732	-3.712760	76.956329	-0.790555	-0.734553	290.590088	-146.717926	141.680267
1	-2.277335	-1.683142	129.638657	-3.107097	78.224899	-0.813110	-0.727224	282.085938	-148.231628	144.015752
2	-2.277430	-1.683072	131.937653	-3.319160	78.702315	-0.813299	-0.727130	285.839355	-149.398315	144.894714

TEST NJMBER 15

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.253712	-1.632620	119.216415	-2.736012	75.966544	-0.858038	-0.658674	266.499268	-143.570572	139.858215
2	-2.255710	-1.636217	126.782898	-4.037299	74.574583	-0.855155	-0.663554	275.218506	-142.957657	137.479645
3	-2.255647	-1.636190	126.974228	-4.015491	74.915283	-0.855128	-0.663519	275.793457	-143.371033	137.922607
1	-2.264157	-1.632447	127.295258	-3.896711	72.668381	-0.868717	-0.658440	271.655518	-139.073196	133.785950
2	-2.264108	-1.632505	128.550812	-4.041291	72.859553	-0.868537	-0.658518	273.655518	-139.621353	134.137924

TEST NUMBER 16

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.277617	-1.739530	114.432480	-1.711743	76.626846	-0.736882	-0.803735	250.151123	-143.396240	141.073659
2	-2.282319	-1.738117	125.216095	-2.136583	81.602990	-0.743500	-0.801819	281.248779	-153.133835	150.234818
3	-2.282465	-1.738032	125.567169	-2.161252	81.693909	-0.743761	-0.801703	281.834473	-153.334869	150.402374
4	-2.282470	-1.738029	125.577835	-2.162268	81.694000	-0.743770	-0.801699	281.847900	-153.336411	150.402542
1	-2.290104	-1.733313	134.228256	-2.530125	84.828064	-0.757805	-0.795299	297.256502	-159.605515	156.172516
2	-2.290143	-1.733235	135.359055	-2.655038	84.863532	-0.757949	-0.795194	298.802002	-159.840485	156.237991

TEST NUMBER 17

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.368742	-1.845277	127.526077	-1.575294	91.039505	-0.584523	-0.947218	299.408936	-169.745499	167.608063
2	-2.369914	-1.847555	148.427856	-1.563622	107.552612	-0.682604	-0.950309	350.530420	-200.131088	198.009491
3	-2.369885	-1.847540	148.562805	-1.553377	107.752838	-0.682597	-0.950288	351.156250	-200.485809	198.378113
1	-2.371849	-1.844112	152.600555	-1.661098	104.624435	-0.589210	-0.945638	349.726807	-194.872437	192.618576
2	-2.371853	-1.844132	152.845306	-1.715092	104.345520	-0.589187	-0.945666	349.604492	-194.432190	192.105072

TEST NUMBER 18

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.497520	-1.738799	147.810928	-3.268583	95.510520	-0.957776	-0.802744	332.520264	-180.274582	175.839600
2	-2.491619	-1.752057	183.652557	-5.129959	111.462921	-0.933887	-0.820732	402.782227	-212.169144	205.208557
3	-2.491198	-1.751154	183.248947	-4.952271	112.031207	-0.934590	-0.819507	403.034668	-213.066345	206.346848
4	-2.491239	-1.751187	183.190125	-4.959242	112.001450	-0.934685	-0.819552	402.947900	-212.928955	206.200012
1	-2.513418	-1.748343	172.492477	-4.156542	102.414337	-0.960725	-0.815692	372.321777	-194.189484	188.549583
2	-2.514222	-1.748564	176.852295	-4.448544	103.388901	-0.961229	-0.815993	380.188721	-197.300430	191.264420
3	-2.514250	-1.748567	177.022797	-4.453040	103.983963	-0.961253	-0.815996	380.546387	-197.481552	191.439438

TEST NUMBER 19

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.189367	-1.731376	119.451111	-1.966081	76.685562	-0.659695	-0.792671	265.968018	-143.849457	141.181778
2	-2.188388	-1.731308	123.314423	-2.166901	78.146072	-0.658808	-0.792580	273.065186	-146.810806	143.870651
3	-2.188404	-1.731310	123.239334	-2.164092	78.109595	-0.658822	-0.792582	272.915527	-146.740005	143.803555
1	-2.194294	-1.726960	119.254791	-1.982204	77.583984	-0.670614	-0.786679	267.469482	-145.525360	142.835815
2	-2.194297	-1.726941	120.036102	-2.073486	77.538803	-0.670643	-0.786654	268.415527	-145.566025	142.752625

TEST NUMBER 20

ITER	F11	F21	PK11	PK12	PK22	P1	P2	P11	P12	P22
0	-2.324999	-1.732354				-0.794000	-0.794000			
0	-2.136999	-1.736776				-0.600000	-0.800000			
1	-2.288272	-1.794724	120.258759	-1.592830	83.357249	-0.672646	-0.878626	273.064209	-155.644302	153.483078
2	-2.289248	-1.795503	132.762085	-1.671544	92.322905	-0.672556	-0.879682	307.253799	-172.238907	169.970871
3	-2.289249	-1.795507	132.850113	-1.669686	92.405121	-0.672560	-0.879689	307.503174	-172.387741	170.122238
1	-2.286997	-1.792812	136.957413	-1.811252	91.795959	-0.673954	-0.876032	310.873291	-171.458328	169.000732
2	-2.287026	-1.792807	136.503967	-1.830860	91.312088	-0.674002	-0.876024	309.582275	-170.594101	168.109909

K-STATISTICS CALCULATED FROM THE ABOVE 20 TESTS

F11	F21	PK21	PK12	PK22	P1	P2	P11	P12	P22
-2.331799	-1.745412	0.011646	-0.000109	0.008150	-0.783082	-0.811721	0.025946	-0.015152	0.015005

REFERENCES

AO-1 Aoki, M., and R.M. Staley, "On Input Signal Synthesis in Parameter Identification," Automatica, vol. 6, pp. 431-440, 197C.

AS-1 Åström, K.J., and P. Eykhoff, "System Identification - A Survey," Automatica, vol. 7, pp. 123-162, 1971.

BA-1 Barankin, E.W., "Locally Best Unbiased Estimates," Ann. Math Stat., vol. 20, pp. 477-501, 1949.

BH-1 Bhattacharyya, A., "On Some Analogues of the Amount of Information and Their Use in Statistical Estimation," Sankhyā, vol. 8, pp. 1-14, 201-218, 315-328, 1946-48.

BR-1 Breakwell, J.V, Stanford University, Stanford, California, personal communications, Spring 1972.

BRY-1 Bryson, A.E., and Y.C. Ho, Applied Optimal Control, Blaisdell Publishing Company, Waltham, Mass., 1969.

DE-1 Denery, D.G., "An Identification Algorithm Which is Insensitive to Initial Parameter Estimates," AIAA Paper 70-34, New York, N.Y., 1970.

DE-2 Denery, D.G., "Identification of System Parameters From Input-Output Data with Application to Air Vehicles," Ph.D. Dissertation, Dept. of Applied Mechanics, Stanford University, Stanford, Calif., 1971.

GI-1 Gilbert, E.G., "Controllability and Observability in Multivariable Control Systems," SIAM J. Control, vol. 1, no. 2, pp. 128-151, 1963.

IC-1 Ichikawa and Tamura, "A New Method Based on Gradient for Solving Optimization Problem with Bang-Bang Control," paper, 1970.

JA-1 Jazwinski, A.H., Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

KA-1 Kailath, T., "An Innovation Approach to Least Squares Estimation, Parts I and II," IEEE Trans. on Automatic Control, pp. 646-661, Dec. 1968.

REFERENCES (Cont)

KAL-1 Kalman, R.E., "Mathematical Description of Linear Dynamical Systems," SIAM J. Control, vol. 1, no. 2, pp. 152-192, 1963.

KAM-1 Kaminski, P.G., at Stanford University, 1971, personal communication.

KR-1 Kroy, W.H., Jr., "Identification of Aircraft States and Parameters," JACC preprint, pp. 302a-302b, 1971.

ME-1 Mehra, R.K., "On the Identification of Variances and Adaptive Kalman Filtering," IEEE Trans. on Automatic Control, pp. 175-184, Apr. 1970.

ME-2 Mehra, R.K., "Identification of Stochastic Linear Dynamic Systems Using Kalman Filter Representation," AIAA J., vol. 9, no. 1, pp. 28-31, Jan. 1971.

ME-3 Mehra, R.K., "Optimal Inputs for Linear System Identification," unpublished paper, 1971.

ME-4 Mehra, R.K., D.E. Stepner, and J.S. Tyler, "A Generalized Method for the Identification of Aircraft Stability and Control Derivatives from Flight Test Date," to be presented at the 1972 JACC, August 16-18, 1972, Stanford University, Stanford, Calif.

ME-5 Mehra, R.K., of Systems Control, Inc., Palo Alto, Calif., personal communication, 1972.

NA-1 Nahi, N.E., and D.E. Wallis, Jr., "Optimal Inputs for Parameter Estimation in Dynamic Systems with White Observation Noise," JACC preprint, pp. 506-513, 1969.

PI-1 Pierre, D.A., Optimization Theory with Applications, John Wiley & Sons, New York, 1969.

SA-1 Sage, A.P., and J.L. Melsa, Estimation Theory with Applications to Communications and Control, McGraw-Hill Book Company, New York, N.Y., 1971.

SA-2 Sage, A.P., and J.L. Melsa, System Identification, Academic Press, New York, N.Y., 1971.

SI-1 Silverman, L.M., "Realization of Linear Dynamical Systems," IEEE Trans. on Automatic Control, pp. 554-567, Dec. 1971.

SP-1 Spain, D.S., "Identification and Modelling of Discrete, Stochastic Linear Systems," Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford University, Stanford, Calif., 1971.

VA-1 Van Trees, H.L., Detection, Estimation, and Modulation Theory, John Wiley and Sons, Inc., New York, N.Y., Part I, 1968, Part II, 1971, Part III, 1971.

WO-1 Wonham, W.M., "Lecture Notes on Stochastic Control," Brown University, Providence, Rhode Island, 1967.

WO-2 Wonham, W.M., "On the Separation Theorem of Stochastic Control," SIAM J. Control, vol. 6, no. 2, 1968.