7-1-1960

A Survey of the Philosophy and State of the Art of Adaptive Systems

G. R. Cooper
Purdue University

Gibson J. E.
Purdue University

Eveleigh V. W.
Purdue University

Lindenlaub J. C.

Meditch J. S.
Purdue University

See next page for additional authors

Authors
G. R. Cooper, Gibson J. E., Eveleigh V. W., Lindenlaub J. C., Meditch J. S., and Raible R. H.

[illegible] Editor File Copy

TR–EE60–1
TECHNICAL REPORT NO. 1
CONTRACT AF 33(616)-6890
PRF 2358

(INTERIM REPORT)
PROJECT 8225
TASK 82181

PURDUE UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

A Survey of the Philosophy and State of the Art of Adaptive Systems

G. R. Cooper, Principal Investigator
J. E. Gibson, Principal Investigator
V. W. Eveleigh
J. C. Lindenlaub
J. S. Meditch
R. H. Raible

Control and Information Systems Laboratory
July 1, 1960
Lafayette, Indiana

FOR
U. S. AIR FORCE
WRIGHT AIR DEVELOPMENT DIVISION
WRIGHT-PATTERSON AIR FORCE BASE
DAYTON, OHIO

Additional copies of this report are available from Engineering Experiment Station, Purdue University, Lafayette, Indiana. Please remit $1.50 to cover the cost of handling.

Make check payable to Purdue University.

TECHNICAL REPORT NO. 1

(INTERIM REPORT)

Contract AF 33(616) - 6890

Project 8225

Task 82181

PRF 2358

A SURVEY OF THE PHILOSOPHY AND STATE OF THE ART OF ADAPTIVE SYSTEMS

for

U. S. AIR FORCE

WRIGHT AIR DEVELOPMENT DIVISION

WRIGHT-PATTERSON AIR FORCE BASE

DAYTON, OHIO

by

G. R. Cooper, Principal Investigator

J. E. Gibson, Principal Investigator

V. W. Eveleigh

J. C. Lindenlaub

J. S. Meditch

R. H. Raible

School of Electrical Engineering

Purdue University

Lafayette, Indiana

July 1, 1960

TABLE OF CONTENTS

LIST OF FIGURES

TABLE OF SYMBOLS

		PAGE
$x_1, x_2, \ldots x_j$	input signals of a control system	7
$y_1, y_2, \ldots y_k$	output signals of a control system	7
p	the operator $\frac{d}{dt}$	8
t	time	8
(t)	a function of time	8
$\leq$	is less than or equal to	8
$\delta(t)$	the unit impulse or Dirac delta function	8
τ	the time at which the unit impulse is applied to the system input	9
(t, τ)	a function of t and τ	9
$<$	is less than	9
$\rightarrow$	approaches	9
∞	infinity	9
s	complex frequency	9
$G(s)$	system transfer function	9
$Y(s)$	Laplace transform of the output signal	9
$X(s)$	Laplace transform of the input signal	9
$g(t)$	impulse response function of a system	12
k	a constant	12

ϵ	the base of natural logarithms	12
$\wp$	damping factor	12
ω	angular frequency (rad./sec.)	12
t_o	time about which g(t) is expanded in a Taylor's series	12
$', '', \ldots {}^{(n)}$	the operators $\frac{d}{dt}, \frac{d^2}{dt^2}, \ldots \frac{d^n}{dt^n}$	12
!	factorial; $n! = n(n-1)(n-2) \ldots 1$	12
ϕ_i	one of a set of orthogonal functions	13
$\neq$	is not equal to	13
σ	real part of s	15
j	imaginary number $\sqrt{-1}$	15
h(t)	impulse response function of a filter	17
w(t)	response of h(t) to the input signal x(t)	17
H(s)	transfer function of the filter	17
$\hat{g}(t)$	approximation to g(t)	18
T	an interval of time	18
$y_s(t)$	response of the system due to stored energy at the initial instant	19
$y_x(t)$	response of the system due to the unit step function applied at the initial instant	19
s_j	a pole of an approximating model	20
$\hat{y}_s(t)$	approximation to $y_s(t)$	20

$0-$	a number, less than 0, but infinitesimal in magnitude	20
$\Delta x(t)$	an increment of $x(t)$	21
t_a	an increment of time	21
$x(n)$	$x(t)$ at $t = nt_a$	21
$y(n)$	$y(t)$ at $t = nt_a$	21
$[x]^{-1}$	inverse of matrix $[x]$	22
$G(z)$	impulse transfer function of a system	23
$Y(z)$	z-transform of the output	23
$X(z)$	z-transform of the input	23
z	z-transform variable	23
$\hat{y}_k(N)$	calculated values of y_k	23
λ	dummy variable of integration	24
$G_m(s)$	transfer function of the model	26
$z_1, z_2, \ldots z_n$	multiplier outputs	28
ϕ_{xy}	cross correlation function	29
$E[\ ]$	expected value	30
ϕ_{xx}	autocorrelation function	30
I.P.	index of performance	38
Mach 2.0	twice the speed of sound	39
$e(t)$	system error	39
F	an arbitrary function	39
$\alpha_1, \alpha_2, \ldots \alpha_n$	desired values of $a_1, a_2, \ldots a_n$	40

$A_1, A_2, \ldots A_n$	arbitrary weighting factors	40
$\bar{c}(t)$	system response vector	40
$\bar{m}(t)$	system control vector	40
$\bar{c}_o(t)$	vector representing the desired state of the system	40
G	cost of deviation function	40
H	cost of control function	40
J	total cost function	40
σ	time	41
q	single process output	41
m	single process input	41
t	present time	42
Q	estimate of desired output	42
M	estimate of desired input	42
λ	arbitrary weighting factor	42
f_q	a function of $Q(\sigma) - q(\sigma)$	42
f_m	a function of $M(\sigma) - m(\sigma)$	42
ISE	Integral of Squared-Error	43
IRAR	Impulse Response Area Ratio	47
IAE	Integral of Absolute Error	48
RMS	Root Mean Square	48
ITAE	Integral of Time multiplied by Absolute Error	48
ITSE	Integral of Time multiplied by Squared Error	49

ISTSE	Integral of Squared Time multiplied by Squared Error	49
ISTAE	Integral of Squared Time multiplied by Absolute Error	49
C(s)	Laplace transform of the output	49
R(s)	Laplace transform of the input	49
ζ_o	damping factor	52
ϵ	error	68
$\ddot{}$	the operator $\frac{d^2}{dt^2}$	70
$\dot{}$	the operator $\frac{d}{dt}$	70
μ	a constant	70
x(t)	system response	70
r(t)	an input function of time	70
A(t)	system matrix function of time	73
y(t)	input function	73

PREFACE

This is a preliminary report on the state of the art of adaptive control. It in no way attempts to review all of the various adaptive systems which have been proposed or constructed. Probably the most complete effort in this direction is WADC TR 59-49, The Proceedings of the Self Adaptive Flight Control Systems Symposium, Edited by P. C. Gregory. Rather this report attempts a synthesis of the present philosophy on adaptive control and is essentially a definition of the problem.

The report attempts to subdivide the adaptive control problem into three subdivisions and to assess present progress in each of these areas. Ideas that have been proposed by various authors are brought together and given unified treatment. In making this organization, various gaps in the present state of the art have become apparent and these are under intensive survey presently at Purdue.

The initial portion of the project, consisting of this organization terminated several months ago and at present the project personnel are engaged on original research along the lines indicated by the monthly progress reports to WADD. Further interim reports will discuss these items and in accordance with present Air Force practice the final report will contain all of the information of the interim reports and will thus be self sufficient.

FOREWORD

This report was prepared by Purdue University, School of Electrical Engineering, Professors G. R. Cooper and J. E. Gibson acting as Principal Investigators, under USAF Contract No. AF 33(616) - 6890, Project No. 8225, Task No. 82181. This contract is administered under the direction of the Flight Control Laboratory, Wright Air Development Division, Wright-Patterson Air Force Base, Ohio, by Lt. P. C. Gregory, the initiator of the study.

CHAPTER I

INTRODUCTION TO THE PHILOSOPHY OF ADAPTIVE CONTROLS

1.0 Abstract

This chapter serves as an introduction to this report by defining terms, discussing the philosophy, and presenting the Purdue viewpoint on adaptive controls. A breakdown of the adaptive process into the functions of identification, decision, and modification is presented. A justification of the adaptive approach to control systems is also given.

1.1 Introduction

This report is concerned with the background material necessary for a study of adaptive controls as well as a summary of the state of the art of adaptive control systems. It is intended that this will lead, in later reports, to detailed work in the various areas that show promise. The first chapter will define some of the terms to be used, discuss the philosophy of adaptive controls, and present the viewpoint of adaptive controls taken at Purdue University.

1.2 What is Meant by an Adaptive Control System?

There is not, as yet, a generally accepted definition of an adaptive control system, but one which has been used here at Purdue is the following: An adaptive system is one which is provided with a means of continuously monitoring its own performance in relation to a given index of performance or optimum condition and a means of modifying its own parameters by closed loop action so as to approach this optimum. This definition implies that an adaptive system must be capable of performing the following functions: provide continuous information about the present state of the system or identify the process; compare present system performance to the desired or optimum performance and make a decision to adapt the system so as to achieve optimum performance; and finally initiate a proper modification so as to drive the control system to the optimum. These three principles, identification, decision, and modification are inherent in any adaptive system. This functional breakdown of an adaptive system is similar to that proposed by Aseltine et. al.[1] Furthermore this breakdown is a useful concept for the design of an adaptive system as it clearly places the adaptive nature in evidence and thus is in agreement with the philosophy of Truxal who states "An adaptive system is any physical system which has been designed with an adaptive viewpoint."[2] A detailed discussion of each of the three phases of the adaptive control problem is presented in the succeeding chapters of this report.

1.3 Justification of the Adaptive Approach

Upon careful consideration of the functions which an adaptive controller must perform, it is obvious that it will be complex in nature and thus may well raise the cost of the overall system by several orders of magnitude. It is, therefore, certainly reasonable to ask why an adaptive control is necessary. A practical problem, which provided early motivation for investigating the adaptive approach, is that of the automatic flight control system. A typical high performance aircraft must be able to operate with satisfactory dynamic response characteristics over a performance envelope which varies from near stall speed to well above the speed of sound, and from air pressures encountered at ground level to the condition of virtually no air at all encountered at very high altitudes. The range of control surface forces required and the variation in dynamic response to be expected over this envelope of performance is extremely large. A possible answer to this problem is the use of a control system which continuously adjusts itself to compensate for these environmental changes, or a self adaptive system. An important advantage of an adaptive system over a control system that is preprogramed to adjust to environmental changes is that it can operate in environments that cannot be predicted from prior knowledge of the flight envelope. This feature is a necessary requirement for exploratory space vehicles intended to operate in unknown environments. Many similar examples, involving missiles, aircraft, chemical processes which are affected by environmental factors, etc. could be cited. Even if no direct applications were immediately apparent, however, a self adaptive system would be of philosophical interest, since it would give greater insight into the problem of designing more "intelligent" systems, systems with a learning capacity.

CHAPTER II

THE IDENTIFICATION PROBLEM

2.0 Abstract

This chapter summarizes the state of the art of the identification problem as it is related to the philosophy of adaptive systems. Requirements of the identification process include the ability to identify the physical system without unduly disturbing its normal operation, and the ability to make the identification in a reasonable amount of time. To date, most of the effort has been devoted to identifying the impulse response or transfer function of linear systems. Representation of an impulse response by means of an analytical function, a graph, sample points, an orthogonal series expansion, and a Taylor's series are discussed. Methods of specifying transfer functions in terms of pole-zero locations, the coefficients of the numerator and denominator polynominals, and by frequency response curves are presented.

Identification techniques proposed in the literature are presented and discussed. The distinction between methods utilizing normal operating signals for identification and those employing a separate test signal is pointed out. Advantages and disadvantages of both approaches are given. The role played by a priori knowledge about the system is discussed. Finally the effect of external noise is mentioned. It is suggested that noise considerations may ultimately determine the choice of an identification technique.

2.1 Introduction - Requirements of an Identification Technique

The general identification problem consists of determining a complete description of the relationships between the input and output signals of an unknown system having input signals $x_1, x_2, \ldots x_j$ and output signals $y_1, y_2, \ldots y_k$. In general the unknown system may be non-linear and time-varying and the number of input signals, j, need not equal the number of output signals, k. The behavior of the unknown system is to be determined by making suitable tests among the various inputs and outputs. This problem has been discussed by Zadeh,[3] Woodrow,[4] Moore,[5] and others.

This chapter, which summarizes the state of the art of the identification problem as it pertains to adaptive control systems, will discuss a somewhat more restricted problem. The system is assumed to be linear or at least the system is to be represented by an equivalent linear system and $j = k = 1$, that is, the system has only one input and one output signal. Note, however, that time-varying systems are not excluded. An identification technique, to be useful in adaptive control systems, must meet two other conditions: first, the identification must be made in the presence of normal operating signals and noise disturbances, and any tests performed upon the system must not unduly disturb the normal operation of the control system; second, the identification must be made relatively quickly if the information is to be useful for the decision-making and modification phases of the adaptive process.

2.2 The Description of Linear Systems

A linear system is one whose input-output characteristics are described by a linear differential equation of the form

$$(a_n p^n + a_{n-1} p^{n-1} + \dots + a_1 p + a_o)x(t) = (b_m p^m + b_{m-1} p^{m-1} + \dots + b_1 p + b_o)y(t) \qquad m \leq n \qquad (2\text{-}1)$$

where x(t) is the input signal, y(t) is the output signal, and p is the operator $\frac{d}{dt}$.* The condition $m \leq n$ is necessary for the physical realizability of the system. In general the coefficients a_i and b_i are functions of time but are independent of x. The behavior of the system is completely determined if all the a_i and b_i are known as functions of time. A useful description of a linear system is the unit impulse response which is the solution of Eq. (2-1) for y(t) when the input signal is a unit impulse, i.e., $x(t) = \delta(t)$ where $\delta(t)$ denotes the delta function.

The theory of linear systems tells us that a knowledge of the impulse response of a linear system gives a complete description of the system. It is possible, by means of the convolution integral, to predict the behavior of the system to any input x(t) if the behavior is known when $x(t) = \delta(t)$.

---

*This chapter will use the notation x(t) or X(s) to represent the input to the "black box" under test and y(t) or Y(s) to represent the output signal. This notation is adopted in order to emphasize the point that the signals used for system identification are in some cases unrelated to the operating signals usually denoted by c and r. The unknown system will be denoted by g(t) or G(s).

The impulse response is denoted by $g(t,\tau)$ and is interpreted as the value of the output at time t when a unit impulse is applied to the input at time τ. When the system is time-invariant $g(t,\tau)$ becomes $g(t-\tau)$ and the impulse response may be represented graphically as in Fig. 2-1. For the time-varying case $g(t,\tau)$ may be represented as the height of a surface above the t, τ plane as shown in Fig. 2-2. It is a property of physical systems that $g(t,\tau) = 0$ for $t < \tau$ (due to the fact that the system cannot respond before the excitation is applied) and $g(t,\tau) \rightarrow 0$ as $(t-\tau) \rightarrow \infty$ (due to losses within the system). Thus, in both Fig. 2-1 and Fig. 2-2, $g(t,\tau)$ is zero for $t < \tau$.

Several of the techniques for system identification require that the system be, at most, slowly time-varying. In terms of Fig. 2-2 this means that variations in the height of the $g(t,\tau)$ surface along lines parallel to the $t = \tau$ line must be slow compared to the significant length of the impulse response.

Taking the Laplace transform of Eq. (2-1) and rearranging terms results in an expression for the system transfer function

$$G(s) = \frac{Y(s)}{X(s)} = \frac{b_m s^m + b_{m-1} s^{m-1} + \ldots + b_1 s + b_o}{a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_o} \tag{2-2}$$

For a time-varying system the coefficients a_i and b_i are functions of time, while for the time-invariant case they are constants. In the intermediate situation, the slowly varying case, the coefficients are essentially constant during the time required to identify the system.

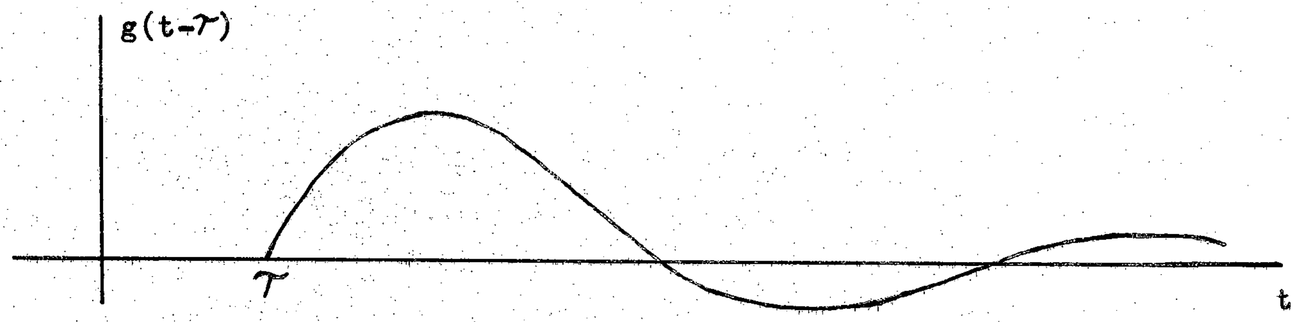

Fig. 2-1

Impulse Response of a Typical
Time-Invariant System

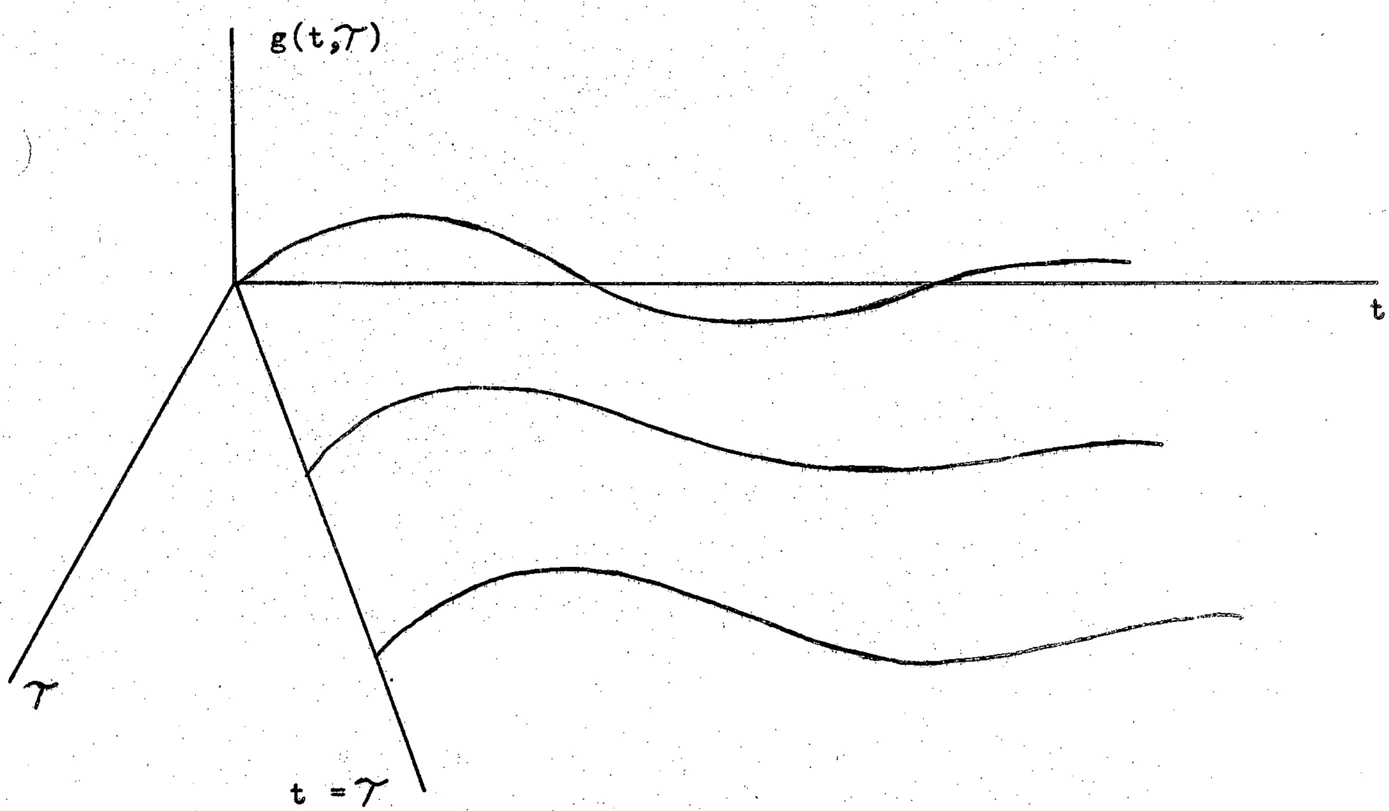

Fig. 2-2

Impulse Response of a Typical Time-Varying System

A graphical representation of the transfer function may be obtained by factoring its numerator and denominator polynominals and plotting the poles and zeros of G(s) on the complex frequency plane. (Fig. 2-3) The poles and zeros are fixed for a time-invariant system and describe a locus for time-varying systems. The movement of the poles and zeros is slow compared to the required measurement time for slowly-varying systems.

Before continuing, it is important to mention that the time-domain description of the system, the impulse response, is entirely equivalent to the frequency domain description, the transfer function. Of the several identification techniques described in section 2.5, some will lend themselves to interpretation in terms of impulse response measurement, and others are more readily described in terms of the transfer function. It is possible, however, to discuss each of the schemes in either the time domain or the frequency domain.

Two ways of describing the behavior of a time-varying linear system have been presented: first, the impulse response $g(t,\tau)$; second, the system transfer function G(s). The impulse response and the transfer function constitute a Laplace transform pair. This fact suggests that additional representations of system behavior could be generated by simply considering other types of transformations. The z transform, commonly used in describing sampled-data systems, is an example. Since nearly all of the work to date has dealt with impulse response measurement or transfer function determination, the discussion in the remainder of this chapter will tend to emphasize these two methods of system description.

Ways of Expressing Impulse Responses

The impulse response of a system may be given as an analytical function of time such as

$$g(t) = k\,\epsilon^{-\zeta\omega t} \sin \omega\sqrt{1-\zeta^2}\; t \qquad (2.3)$$

for a simple second-order system. Another common way of representing an impulse response is by means of a graph such as the one in Fig. 2-1 (or Fig. 2-2 for the time-varying case). Sometimes, instead of a complete graph, only sample points of the impulse response curve are given. (Fig. 2-4) In practice some error is introduced by the sampling process, but in most engineering applications this error approaches zero as the number of sampling points approaches infinity.

Another method of representing an impulse response is by a Taylor's series expansion.

$$g(t) = g(t_o)+(t-t_o)g'(t_o)+\frac{(t-t_o)^2}{2!}\,g''(t_o)+\;\ldots\;+\frac{(t-t_o)^n}{n!}\,g^{(n)}(t_o)+\ldots \qquad (2-4)$$

The nature of impulse responses of practical systems indicates that, in general, a large number of terms will be required in the Taylor's series expansion to achieve a good approximation to the actual impulse response.

A different kind of series representation of the impulse response, also useful in the identification problem, is a series of orthogonal functions. g(t) may be expressed as

$$g(t) = \sum_{n=1}^{\infty} c_i \, \phi_i \, (t) \tag{2-5}$$

where the ϕ_i's are a set of orthogonal functions satisfying the conditions

$$\int_0^{\infty} \phi_i \, \phi_j \, dt = 0 \qquad i \neq j \qquad i,j = 1,2,3,\ldots \tag{2-6}$$

and

$$\int_0^{\infty} \phi_i^{\,2} \, dt \neq 0 \qquad i = 1,2,3,\ldots \tag{2-7}$$

and the constants, c_i, are given by

$$c_i = \int_0^{\infty} g(t) \, \phi_i(t) \, dt \tag{2-8}$$

When the integral in Eq. (2-7) equals 1 for all i, the set of functions is said to be orthonormal. If the set of orthogonal functions, ϕ_i, can be choosen properly the series, (2-5), will converge rapidly.

Ways of Specifying Transfer Functions

The expression of a transfer function as a ratio of two polynominals in s was given in Eq. (2-2). If the coefficients of these polynominals are known as functions of time the transfer function is completely specified.

A common graphical representation of a transfer function, the pole-zero plot, an example of which is shown in Fig. 2-3, is an alternate way of specifying a system's transfer function. A method of specification, which is closely related to the specification of pole-zero locations, is a knowledge of the order and location of the poles of the transfer function and the residue associated with each pole.

Often the form of the transfer function is not known; in these situations a very useful graphical method of specifying a transfer function is by means of a frequency response curve or Bode plot. An example of a typical magnitude and phase plot is shown in Fig. 2-5. Very often it is not necessary to specify the complete frequency response curve and only sample points on the curves are obtained.

The Laplace transform of Eq. (2-5) would result in an expansion of the system transfer function in a series of orthogonal complex functions. The coefficients of this series may be used to specify the transfer function in the same manner as the coefficients of Eq. (2-5) are used to specify the system impulse response.

This summary of methods of representing the impulse response and transfer function of a linear system will serve as a background for the discussion of the various identification techniques that have been proposed in the literature.

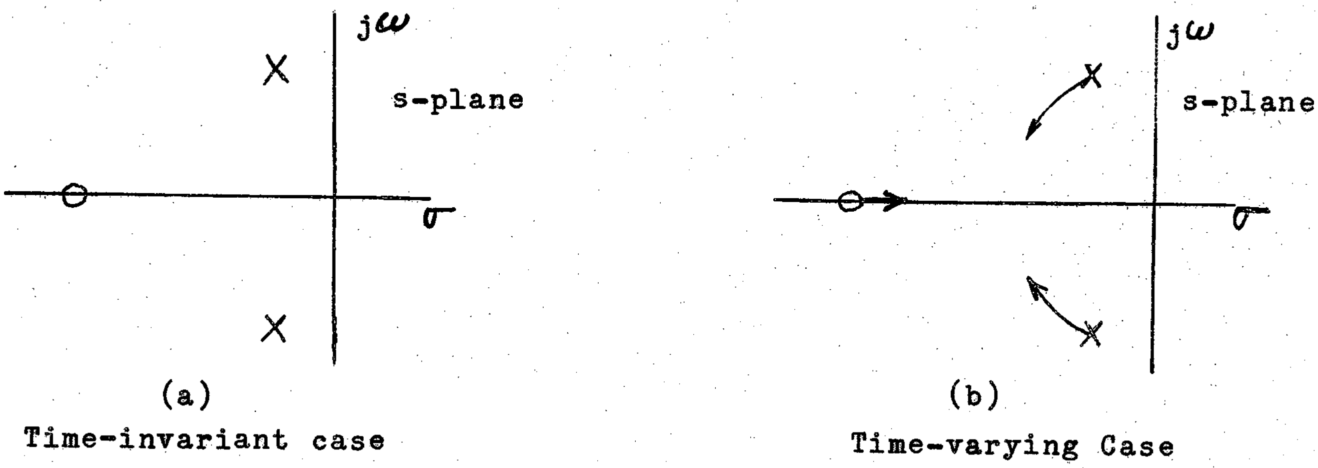

Fig. 2-3

Typical Pole-Zero Plots

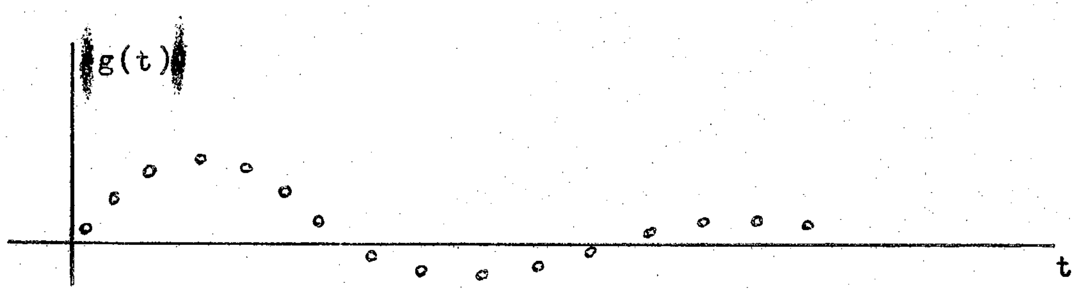

Fig. 2-4

Sample Points of an Impulse Response

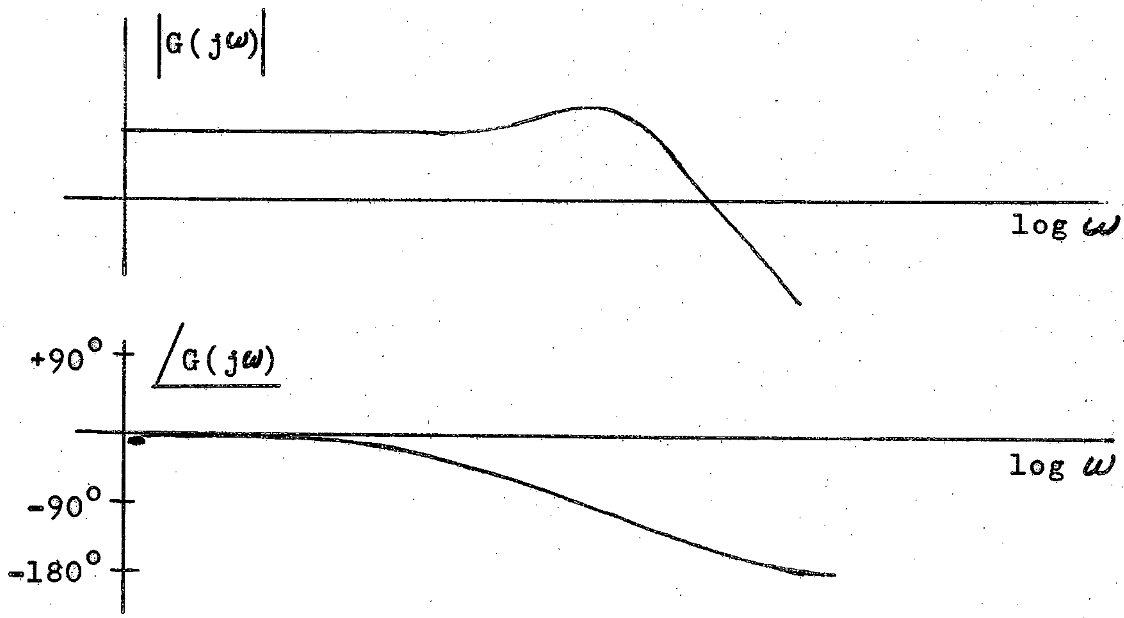

Fig. 2-5

An Example of Magnitude and Phase Plots

2.3 Proposed Identification Techniques

The purpose of this section is to summarize the various techniques that have been proposed for solving the identification problem in adaptive systems. Some of the techniques described have been developed with the adaptive control problem in mind, while others have considered the identification problem on its own merit. The discussion here will be from the adaptive control system viewpoint. All of the proposed methods are applicable to time-invariant or slowly-varying linear systems.

The basis for determining the impulse response of a linear system lies in the convolution integral

$$y(t) = \int_{-\infty}^{t} x(\tau)\, g(t,\tau)\, d\tau \qquad (2\text{-}9)$$

where x(t) and y(t) are the input and output signals respectively and $g(t,\tau)$ is the impulse response which describes the system. For the time-invariant case, Eq. (2-9) becomes

$$y(t) = \int_{-\infty}^{t} x(\tau)\, g(t-\tau)\, d\tau \qquad (2\text{-}10a)$$

or, upon a change in variable

$$y(t) = \int_{0}^{\infty} g(\tau)\, x(t-\tau)\, d\tau \qquad (2\text{-}10b)$$

This second formulation of the convolution integral is approximately correct for slowly-varying systems; i.e., systems whose parameters do not change appreciably during the time required to measure the impulse response.

A direct graphical representation of the impulse response could be obtained by simply applying an impulse at the input and observing the output signal. From a practical standpoint this scheme is unsuitable because in many cases an impulse applied at the input of the control system would seriously disrupt the normal operation of the system. Also, a system that can be adequately represented by a linear system for normal operation may often exhibit nonlinear characteristics for large input signals, such as impulses.

A special case of a more general result proposed by Turin[6] is another way of obtaining a direct graphical representation of g(t). Fig. 2-6 illustrates Turin's method. A known signal, x(t), is applied to the input of the system under test and the output signal is passed through a filter h(t). h(t) is designed so that its output signal is an estimate of g(t). The determination of h(t) is easily illustrated if the order of g(t) and h(t) is reversed. (Fig. 2-7) For linear systems the order of operations is unimportant. From Fig. 2-7 it is apparent that the signal w(t) must be an impulse if the signal on the far right is to equal g(t). Thus the function of h(t) is to convert the signal x(t) into an impulse. If the effects of noise in the system are neglected, the transfer function of h(t) is given by

$$H(s) = \frac{1}{X(s)} \tag{2-11}$$

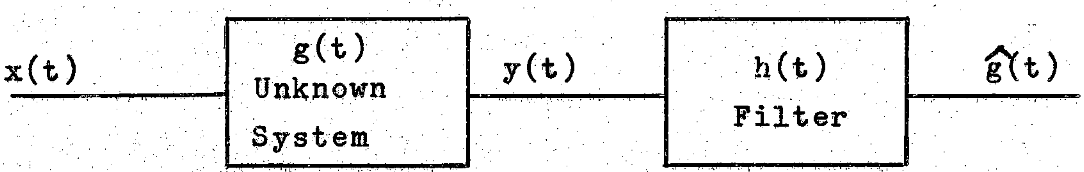

Fig. 2-6

Method of Measuring g(t) Proposed by Turin

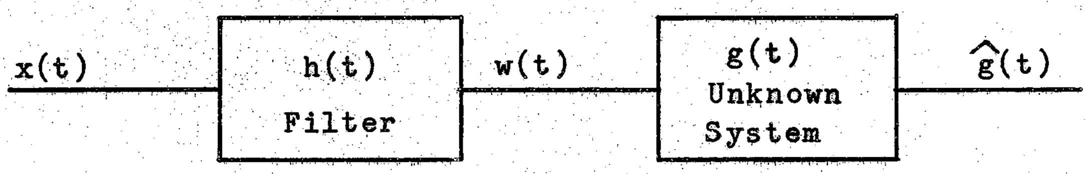

Fig. 2-7

g(t) and h(t) Interchanged

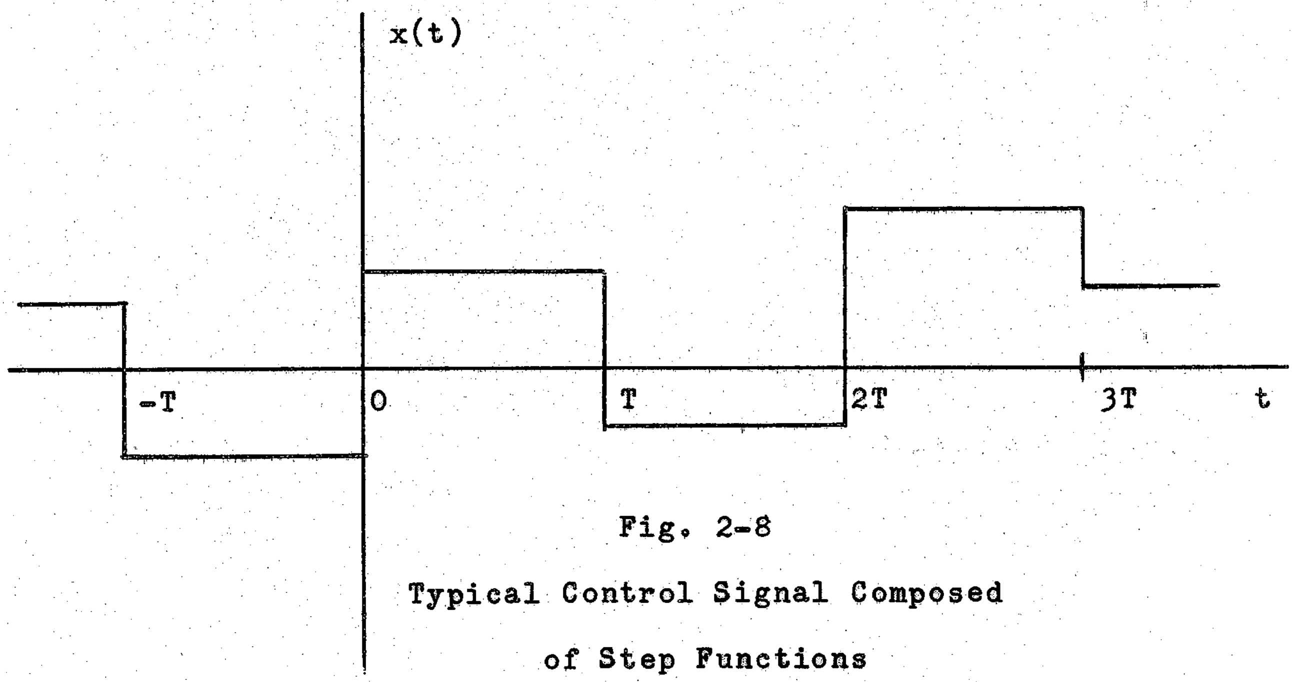

Fig. 2-8

Typical Control Signal Composed

of Step Functions

The advantage of this technique is that, by using the arrangement of Fig. 2-6, one can obtain a direct graphical presentation of the system's impulse response without the necessity of applying an impulse to the system.

Turin's method of impulse response measurement is most easily used in interval control adaptive systems employing a fixed form of input signal. If the form of the control signal is not fixed the filter h(t) must vary with time in such a manner that H(s) is approximately equal to 1/X(s).

An example of an interval control system employing a control signal composed of step functions has been suggested by Braun.[7] A typical input signal of this type is shown in Fig. 2-8. Consider the control interval beginning at t = 0. The convolution integral relating the input and output of the system is divided into two intervals,

$$y(t) = \int_{-\infty}^{0} x(\tau)\, g(t-\tau)\, d\tau + \int_{0}^{t} x(\tau)\, g(t-\tau)\, d\tau \qquad (2\text{-}12)$$

$$= y_s(t) \quad + \quad y_x(t)$$

where the first term represents the response of the system (for $t > 0$) due to its initial stored energy and the second term results from the step applied at t = 0. $y_x(t)$ cannot be measured directly but must be computed from

$$y_x(t) = y(t) - y_s(t) \qquad (2\text{-}13)$$

Hence it is necessary to determine the stored energy term $y_s(t)$. This is achieved in Braun's procedure by, so far as the stored energy term is concerned, approximating the unknown system by a model with fixed pole positions, s_j. An approximation to the stored energy term is then given by

$$y_s(t) \approx \hat{y}_s(t) = \sum_{j=1}^{n} \alpha_j \, \epsilon^{s_j t} \qquad (2\text{-}14)$$

The α_j's are picked so that $y_s(t)$ is a good approximation of $y_s(t)$. This is achieved by choosing the α_j's so that the first p derivatives of $\hat{y}_s(t)$ equal the first p derivatives of $y_s(t)$. From Eq. (2-12), since $y_s(t)$ is continuous at t = 0, it follows that

$$y_s(0) = y(0-)$$
$$y_s'(0) = y'(0-)$$
$$\qquad (2\text{-}15)$$
$$y_s^{(p)}(0) = y^{(p)}(0-)$$

Thus by measuring the first p derivatives of the output signal at t = 0- and setting $\hat{y}_s(0) = y(0-)$, $\hat{y}_s'(0) = y'(0-)$, . . ., $\hat{y}_s^{(p)}(0) = y^{(p)}(0-)$ one is able to obtain an approximation to $y_s(t)$ and hence to $y_x(t)$, the system step response.

Braun proposed, at this point, to apply $y_x(t)$ to the input of an orthogonal spectrum analyzer and use the coefficients of the resulting orthogonal series to identify the system. Alternatively, following Turin's idea, the step response can be differentiated (H(s) = s) to obtain a direct representation of the system impulse response.

The expansion of the impulse response in a Taylor's series, also suggested by Braun,[8] constitutes another means of representing the system impulse response. Braun shows that one can compute a Taylor's series expansion about a point t_0 by applying an abrupt change in the input signal, $\Delta x(t)$, at t_0 and measuring the derivatives of the output signal just prior to, and just after $t = t_0$. In the example discussed by Braun $\Delta x(t)$, the change in control signal applied at $t = t_0$, is a sum of singularity functions.

It is felt that here, as well as in Braun's orthogonal expansion technique, the necessity of measuring the derivatives of the output signal imposes a serious practical limitation on the method.

The application of an impulse to the input of $g(t)$, or the technique suggested by Turin, avoids the necessity of solving the integral equation Eq. (2-10). An alternate procedure is to apply a known signal to the input, measure the output signal, and actually solve Eq. (2-10). In general the exact solution is very difficult. A way of circumventing this difficulty has been suggested by Levine,[9] Woodrow,[10] and Cooper.[11] Instead of a continuous description of $x(t)$, $y(t)$, and $g(t)$, these quantities are represented by sample points spaced t_a seconds apart. $x(t)$ and $y(t)$ are denoted by $x(n)$ and $y(n)$, respectively, where the n indicates the nth sampling point corresponding to the instant of time $t = nt_a$, and $g(t)$ is denoted by g(p) and is assumed to have the property g(p) = 0 when $p < 0$ and also when $p > P$ for some $P > 0$ where P is some positive integer. x and y are observed for N sample periods. The results are more easily expressed if the following matrix notation is adopted.

$$[y] = \begin{bmatrix} y(P) \\ y(P+1) \\ \cdot \\ \cdot \\ y(P+N) \end{bmatrix} \qquad [g] = \begin{bmatrix} g(0) \\ g(1) \\ \cdot \\ \cdot \\ g(P) \end{bmatrix}$$

$$[x] = \begin{bmatrix} x(P) & x(P-1) & \cdot\ \cdot\ \cdot & x(0) \\ x(P+1) & x(P) & & x(1) \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ x(P+N) & x(P+N-1) & \cdot\ \cdot\ \cdot & x(N) \end{bmatrix}$$

Then

$$[y] = [x]\ [g] \tag{2-16}$$

and

$$[g] = [x]^{-1}\ [y] \tag{2-17}$$

where $[x]^{-1}$ is the inverse of the matrix $[x]$. To insure the linear independence of these equations it is necessary and sufficient that the x(n) sequence not be the solution of any linear difference equation of order P or lower for $0 \leq n \leq N + P$.

Sampling the operating signals at the input and output of a control system, forming the appropriate $[x]^{-1}$ and $[y]$ matrices, and using Eq. (2-17) is a means of identifying the characteristics of a linear adaptive control system.

An identification technique which employs samples of the input and output signals as well as a mathematical model of the unknown system has been proposed by Kalman.[12] Kalman chooses to describe the unknown system in terms of the pulse transfer function, $G(z)$.

$$G(z) = \frac{Y(z)}{X(z)} = \frac{a_o + a_1 z^{-1} + \ldots + a_n z^{-n}}{1 + b_1 z^{-1} + \ldots + b_n z^{-n}} \qquad (2\text{-}18)$$

The number n corresponds to the order of the system and is determined either from a priori knowledge of the system's order or an engineering decision to represent the true system by an n^{th} order approximation.

The input and output are related by the difference equation

$$y_k + b_1 y_{k-1} + \ldots + b_n y_{k-n} = a_1 x_{k-1} + a_2 x_{k-2} + \ldots + a_n x_{k-n} \qquad (2\text{-}19)$$

which can be solved for y_k yielding

$$y_k = a_1 x_{k-1} + a_2 x_{k-2} + \ldots + a_n x_{k-n} - b_1 y_{k-1} - \ldots - b_n y_{k-n} \qquad (2\text{-}20)$$

At the N^{th} sampling instant the coefficients will be denoted by $a_i(N)$ and $b_i(N)$. This set of coefficients and Eq. (2-20) can be used to calculate all past values of y_k. The y_k's computed in this way will be called $\hat{y}_k(N)$'s and are given by

$$\hat{y}_k(N) = a_1(N)x_{k-1} + a_2(N)x_{k-2} + \ldots + a_n(N)x_{k-n} - b_1(N)y_{k-1} - \ldots - b_n(N)y_{k-n}$$

$$\text{for } k = 0, 1, 2, \ldots, N \qquad (2\text{-}21)$$

The $a_i(N)$'s and $b_i(N)$'s in this equation are chosen so that the mean square difference between the measured y_k's and the computed

$\hat{y}_k(N)$'s is minimized, that is the $a_i(N)$'s and $b_i(N)$'s are picked in a manner that will minimize the expression

$$\frac{1}{N} \sum_{k=o}^{N} \left[y_k - \hat{y}_k(N) \right]^2 \tag{2-22}$$

The $a_i(N)$'s and $b_i(N)$'s chosen in this manner serve to identify the system at the N^{th} sampling instant. The procedure is repeated at each successive sampling instant.

Corbin[13] has proposed a method of continuously measuring the location of the poles, the zeros, and the gain factor of a system transfer function by analog techniques. The procedure will be illustrated for a first order system

$$G(s) = \frac{K}{s + a} \tag{2-23}$$

where K and a are unknown. The corresponding differential equation is

$$\frac{dy(t)}{dt} + a\, y(t) = K\, x(t) \tag{2-24}$$

Upon integration and solution for K and a

$$K(t) = \frac{a \int_o^t y(\lambda)\, d\lambda + y(t) - y(0)}{\int_o^t x(\lambda)\, d\lambda} \tag{2-25}$$

$$a(t) = \frac{k \int_o^t x(\lambda)\, d\lambda + y(0) - y(t)}{\int_o^t y(\lambda)\, d\lambda} \tag{2-26}$$

Eq. (2-25) and Eq. (2-26) show that, even in the first-order case, there is cross-coupling among the unknown constants, and, in addition, the initial conditions of x(t) and y(t) must be known. This is a particular drawback in the higher order cases because it would require measuring the derivatives of x(t) and y(t) at the beginning of each computation period.

A number of identification techniques which employ a model of the physical system have been suggested. Margolis and Leondes[14] propose the use of a "learning model" for system identification and Whitaker et. al.[15] discuss an adaptive flight control system employing a model that has been built and flight tested by an M.I.T. research group. The general approach using the model technique is the following: if the order of the system to be measured is known, a model of the same order is chosen; if the order is not known, the engineer decides to represent the unknown system by an n^{th} order system where n is based upon some a priori knowledge about the system and perhaps a certain amount of engineering judgment. A block diagram showing an identification technique employing a model is proposed in Fig. 2-9. The difference between the output of the system under test and the output of the model is a measure of the degree of "goodness" for the model. When the model is an exact replica of the unknown system the error signal will be zero. A parameter adjustment computer adjusts the parameters of the model until some function of the error signal is satisfied.[15] The nature of the parameter adjustment computer varies with the application.

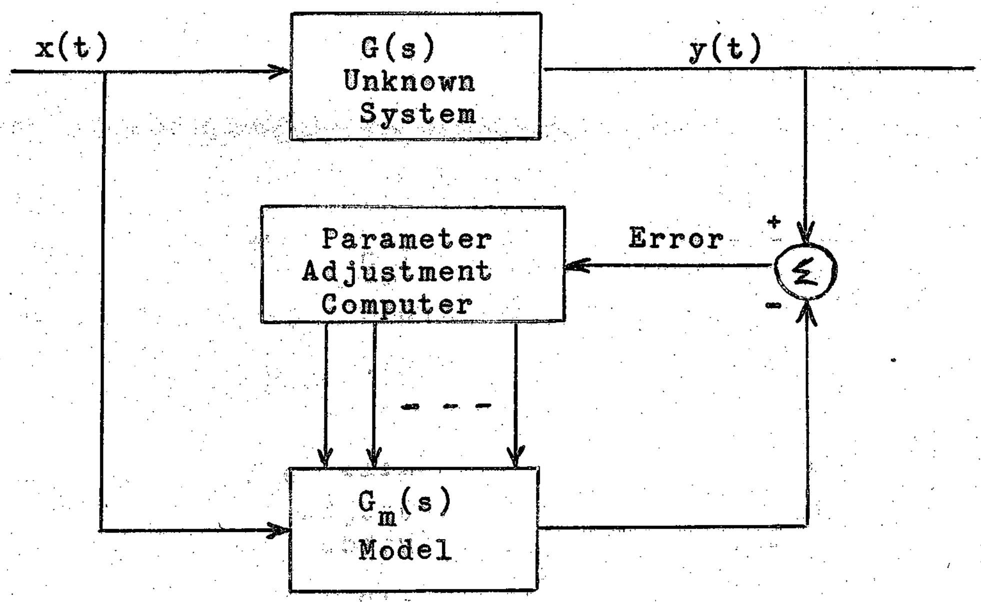

Fig. 2-9

Identification by Means of a Model

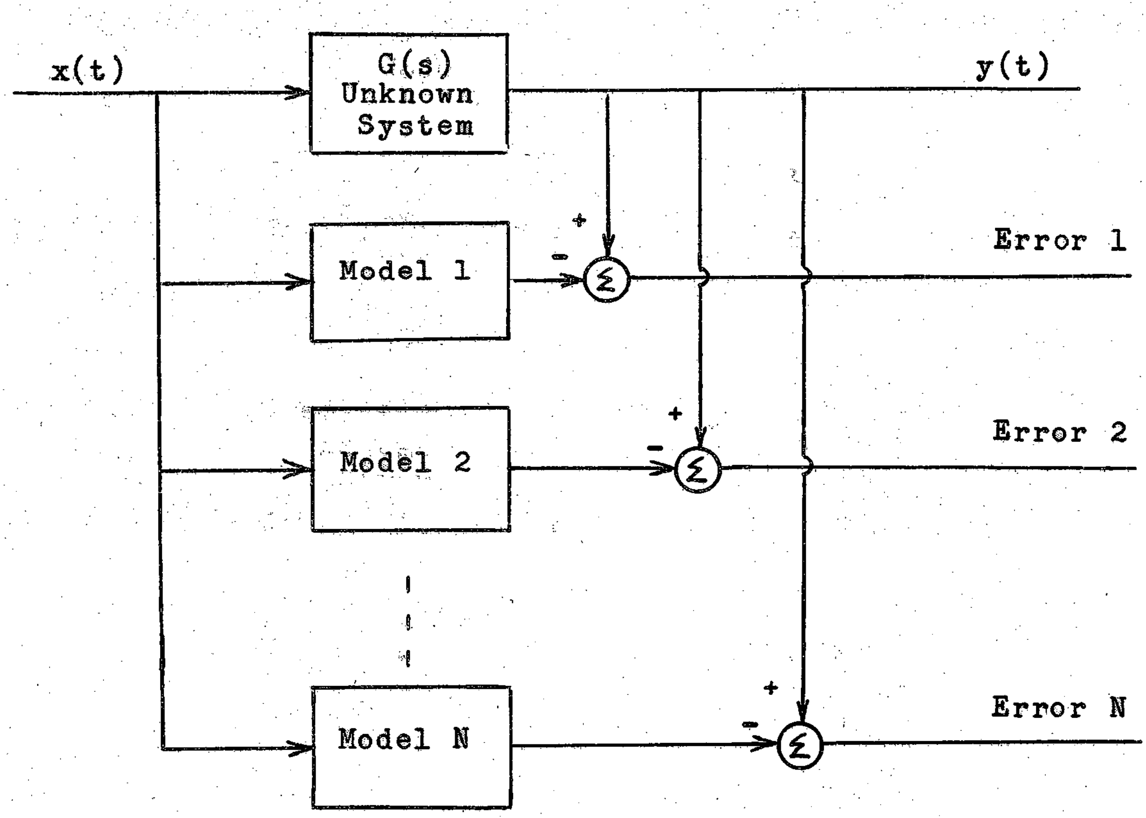

Fig. 2-10

Identification Using N Models

Fig. 2-9 is drawn from the identification problem viewpoint and the diagram emphasizes the identification operation by showing adjustment of the parameters of the model to track the system under test. In an actual adaptive control system the model may represent the optimum system and the computer may adjust several control system parameters in such a way as to force the control system to follow the model. In such a system one cannot completely isolate the aspects of identification, decision, and modification for they are carried out simultaneously.

An alternate model approach, useful if the range of parameter variations is known, is shown in Fig. 2-10. N models are used and the channel with the smallest error at any particular time is chosen to represent the system. If the range of parameter variations is large, or an accurate description of the unknown is desired, the number of models will be large, while if the control system is known to belong to a limited class the number of models will be small.

The identification of a system can be achieved only if energy is supplied to the system and the response observed. In all of the identification techniques mentioned above, the energy is supplied by the normal operating signals. A particular advantage of using the operating signals for process identification is that it is not necessary to disturb the normal performance of the system; a disadvantage is that if, in the course of normal operation, the input signal is identically zero for any appreciable length of time, no information about the system behavior can be obtained. These methods of identification are therefore limited to those cases where the input signal is never zero for an appreciable length of time.

An identification technique employing crosscorrelation which does not depend upon the operating signal has been proposed by Anderson et. al.[16,17] A noise signal, whose amplitude is small compared to normal control signals, is applied to the input of the system. The output signal is then crosscorrelated with the input. When the noise is suitably chosen, the input-output crosscorrelation function has the same form as the impulse response of the system under test. A block diagram of a crosscorrelator is shown in Fig. 2-11.

The output of g(t) is related to the input by the convolution integral

$$y(t) = \int_{-\infty}^{+\infty} x(t-\lambda_1)\, g(\lambda_1)\, d\lambda_1 \qquad (2\text{-}27)$$

The output of the multiplier $y(t)\, x(t-\tau_i)$ is given by

$$z_i(t) = \int_{-\infty}^{+\infty} x(t-\tau_i)\, x(t-\lambda_1)\, g(\lambda_1)\, d\lambda_1 \qquad (2\text{-}28)$$

The smoothing filter has the effect of taking the average value of z(t)

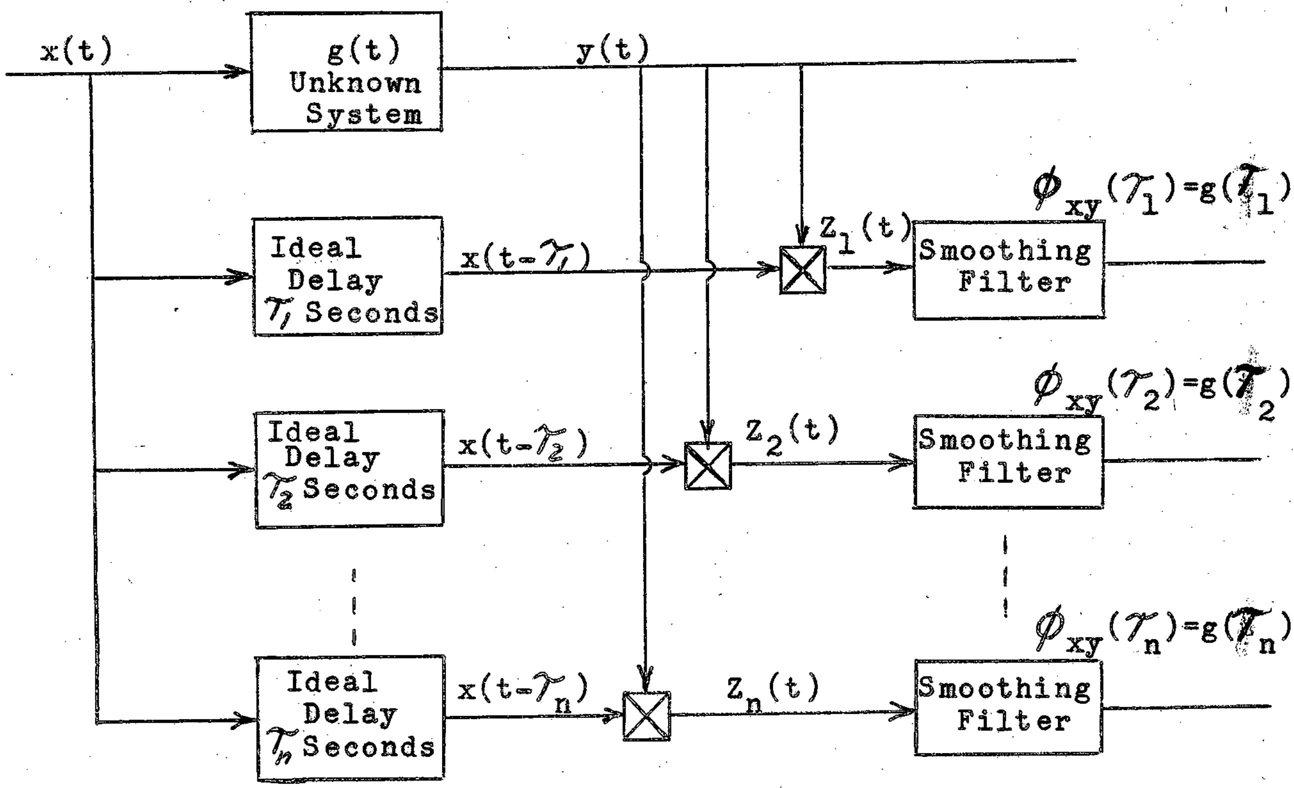

Fig. 2-11

Identification by Crosscorrelation

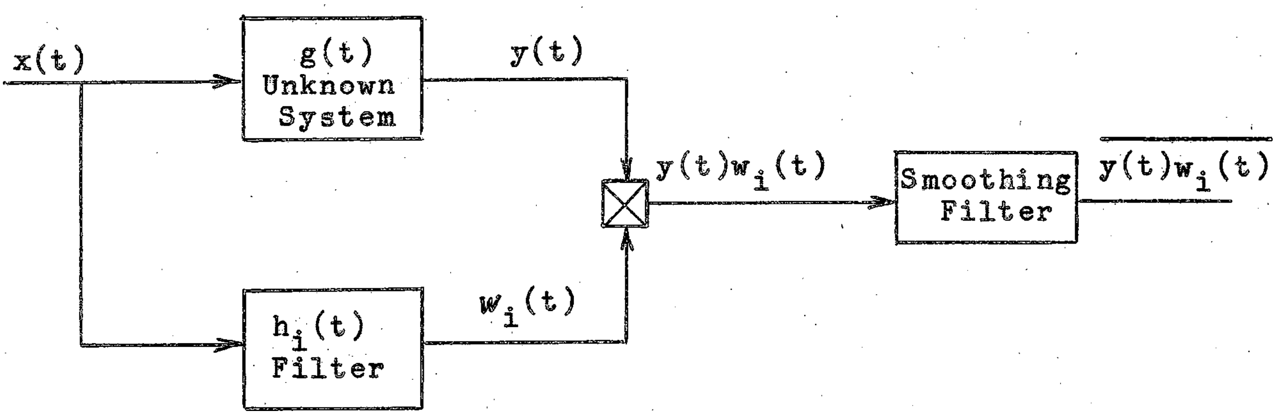

Fig. 2-12

Single Channel of an Orthogonal Spectrum Analyzer

$$z_i(t) = \phi_{xy}(\tau_i) = E\left[\int_{-\infty}^{+\infty} x(t-\tau_i)\, x(t-\lambda_1)\, g(\lambda_1)\, d\lambda_1\right]$$

$$= \int_{-\infty}^{+\infty} E\left[x(t-\tau_i)\, x(t-\lambda_1)\right] g(\lambda_1)\, d\lambda_1 \qquad (2\text{-}29)$$

$E\left[x(t-\tau_i)\, x(t-\lambda_1)\right]$ is recognized as the input noise autocorrelation function $\phi_{xx}(\tau_i - \lambda_1)$. If $x(t)$ is assumed to be white noise $\phi_{xx}(\tau) = \delta(\tau)$ and

$$\phi_{xy}(\tau_i) = g(\tau_i) \qquad (2\text{-}30)$$

In practice, it is never possible to generate white noise, but if the bandwidth of the noise is wide compared to the bandwidth of $g(t)$, Eq. (2-30) is approximately correct. Anderson, et. al.(16) discuss the use of discrete interval binary noise as an input test signal. Each channel of the correlator shown in Fig. 2-11 furnishes one sample point on the impulse response.

A variation of the correlation technique, suggested by Cooper* is illustrated in Fig. 2-12. The ideal delay is replaced by a filter with impulse response $h_i(t)$. $y(t)$ is again given by Eq. (2-27) and

$$w_i(t) = \int_{-\infty}^{\infty} x(t-\lambda_2)\, h_i(\lambda_2)\, d\lambda_2 \qquad (2\text{-}31)$$

The average value of the multiplier output is

$$\overline{w_i(t)\, y(t)} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} E\left[x(t-\lambda_2)x(t-\lambda_1)\right] h_i(\lambda_2) g(\lambda_1) d\lambda_1 d\lambda_2 \qquad (2\text{-}32)$$

---

*Personal discussion with Dr. G. R. Cooper, May, 1960.

Assuming a white noise input

$$\overline{w_i(t)\ y(t)} = \int_{-\infty}^{+\infty} g(\lambda_1) h_i(\lambda_1)\ d\lambda_1 \qquad (2-33)$$

Comparing Eq. (2-33) and Eq. (2-8) shows that if $h_i(t)$ is made equal to $\phi_i(t)$, the arrangement of Fig. 2-12 can be used to measure the coefficients of an orthogonal function series expansion of g(t). Fig. 2-12 is, in effect, a single channel of an orthogonal spectrum analyzer.

A spectrum analyzer with a set of filters, $h_i(t)$, having transfer functions $\sin \omega_1 t$, $\cos \omega_1 t$, $\sin \omega_2 t$, $\cos \omega_2 t$, . . . $\sin \omega_n t$, $\cos \omega_n t$, . . . could be used to obtain sample points for the curves of $|G(j\omega)|$ vs ω and $\angle G(j\omega)$ vs ω.

The advantage of a correlation type identification scheme is that the measurement does not depend upon the presence of a control signal. The correlator outputs furnish continuous information about g(t) even if the input control signal is zero. Also, the correlator outputs are not effected by the presence of normal input signals as long as the test signals, x(t), and the normal inputs are statistically independent.

2.4 Summary and Conclusions

The methods of identifying the characteristics of a slowly-varying linear system that have been discussed in the last section may be grouped in two ways; methods that depend upon the normal input control signal for system identification as compared to methods that make use of a test signal for system identification, or methods that require no a priori knowledge of the system vs. methods that require a knowledge of the form or order of the system.

The organization of the material in section 2.3 followed the first type of classification. All of the methods proposed, except those employing correlation techniques, used energy supplied by the normal operating signals to identify the system. These methods are not satisfactory if the control system is of such a nature that the input is zero for appreciable lengths of time. The correlation techniques employ a noise type test signal and do not rely on control signals to supply the energy necessary for identification.

The identification methods suggested by Kalman, and Corbin, as well as the model techniques, all required a knowledge of at least the order of the system. All of the other techniques required no a priori knowledge of the system.

The first of the two basic requirements of an identification scheme mentioned at the beginning of this chapter was that the identification process must not disturb the normal operation of the system. Each of the methods mentioned above satisfies this requirement. No mention has been made of the ability of the various techniques to satisfy the second requirement, that of making the identification in a relatively short amount of time. Each of the schemes discussed requires a certain minimum time to measure, with a specified accuracy, the characteristics of the unknown system, and this minimum measurement time is closely related to the effects of external noise upon the system. While Anderson et. al.[16] discuss the minimum smoothing time required in the crosscorrelator, and Levine[9] gives an expression for the variance of the estimate of the impulse response sample points, the problem of the effects of external noise upon the various identification techniques has been, for the most part, ignored in the literature.

The effects of external noise upon the identification of a linear system is currently under study at Purdue University. While no detailed conclusions can be drawn at this time, the following statements can be made. The identification time is inversely proportional to the accuracy demanded of the measurement; i.e. greater accuracy must be paid for by longer identification time. Also, the measurement time is inversely related to the a priori knowledge of the system; i.e. the greater the a priori knowledge, the shorter the measurement time required. As an example, consider the limiting case where the system is known exactly. Then it is not even necessary to make a measurement to identify the system . . . identification can be achieved in zero time. The nature of the relationship between measurement time and a priori knowledge is not known at present but the example just mentioned does illustrate the point that the ultimate rapid identification scheme must make use of all available a priori knowledge about the system.

The effects of noise cannot be ignored by the engineer and should be one of the factors influencing the choice of an identification procedure for any particular application. The methods suggested by Braun and Corbin, which require the measurement of higher order derivatives, will be particularly susceptible to noise problems. Because of this, these methods are not deemed practical for most situations. All of the other identification techniques will work in the presence of noise, but a critical comparison of the various methods cannot be made at this time.

The aim of the identification problem, as presented in this chapter and as presented in the literature to date, has been to obtain a complete description of the input-output relationships of a linear system. A very important and basic question arises at this point. Is a complete description of the adaptive control system necessary? True, knowing the impulse response or transfer function enables the engineer to compute any other properties of the system he might desire. Perhaps, however, it would be easier and faster to measure these other quantities directly. Anderson et. al.[16] have suggested that the three quantities, gain, rise time, and overshoot, might serve to describe a system in so far as adaptive control is concerned. Is it easier and faster to measure these quantities directly, or is there an advantage to calculating these quantities from a knowledge of the transfer function or impulse response? These and other questions relating to the fundamental nature of adaptive control systems provide sufficient motivation for continued research in the area.

CHAPTER III

THE DECISION PROBLEM

3.0 Abstract

The decision problem deals with the development and specification of analytical methods by which system performance can be evaluated and from which a strategy to achieve adaptation can be evolved. The most common method of system evaluation is the use of an index of performance which is defined as a functional relationship involving system characteristics in such a manner that the optimum operating characteristics may be determined from it.

In this chapter a number of indices of performance are reviewed proceeding from a general formulation to particular cases which have been treated in detail in the literature. A review of the literature is presented to establish the present status of the decision problem. Finally, three important limitations of indices of performance are discussed: usefulness, uniqueness, and selectivity.

3.1 Introduction

Once the dynamic elements to be controlled have been identified or characterized, a more or less complex decision process must be involved in deciding how to readjust the system.(18) This decision process involves an index of performance* to which the present performance must be compared in order to evolve a plan of action. The physical means for evaluating the index of performance may be instrumented into the adaptive loop directly or the system may be required to develop its own index of performance by a goal-seeking or learning process. Hence, the decision problem is concerned with the development and specification of analytical methods by which system performance can be evaluated and from which a strategy to achieve adaptation can be evolved. In order to be of any practical significance, of course, such methods must be capable of instrumentation.

---

*An index of performance is defined as a functional relationship involving system characteristics in such a manner that the optimum operating characteristics may be determined from it.

3.2 The Index of Performance

The notion of an index of performance has already been defined, but before elaborating on some of the work that has been done on the decision problem, it will be worth while to comment briefly on certain ideas which underlie the formulation of an index of performance.

The purpose of an index of performance (hereafter abbreviated I.P.) is to define the present state of the dynamic process or elements with respect to an optimum state thereby supplying information which indicates where the dynamic process is with respect to this optimum. Moreover, it should indicate what must be done to achieve the optimum state. The I.P. may be either a minimum, a maximum, a null, or simply a particular number at the optimum. While analytically tractable, a large number of I.P.'s are impossible to construct during normal system operation without disturbing the system excessively. Hence, in designing the decision portion of the adaptive loop, the engineer may have to accept an inferior I.P. which does not supply all the information required but is easy to measure in preference to an ideal I.P. which supplies all the information required but is impossible to measure experimentally.

The use of only one I.P. is a rather common practice in present day analytical design theory.[19] However, because of the inherent complexity of adaptive control systems, it is felt that more than one I.P. will be required for purposes of evaluating these systems. For example the use of a given I.P. may be very adequate to evaluate the dynamics of a supersonic aircraft near stall conditions but totally useless to give a measure of the aircraft's dynamics under level flight at Mach 2.0. Under such circumstances a proper weighting of a number of I.P.'s as a function of environment will be necessary to evaluate the aircraft's characteristics over the entire range of its flight envelope. In essence, this is goal-seeking as mentioned in Section 3.1. The system changes the I.P. or switches from one I.P. to another in accordance with some higher goal.

The most common I.P.'s used are those which employ some arbitrary function of system error. System error is defined to be the difference between the desired value of the system response and the actual value of the system response. Symbolically,

$$\text{I.P.} = F\left[e(t)\right] \tag{3-1}$$

where

$e(t)$ = system error as a function of time.

F = some arbitrary functional operation.

The extension of Eq. (3-1) to the multi-dimensional case follows by considering a number of error signals representative of the same number of system aspects which are to be controlled. For this case the I.P. becomes

$$\text{I.P.} = F\left[e_1(t),\ e_2(t),\ \ldots\ ,\ e_N(t)\right] \tag{3-2}$$

where

$e_i(t)$ = the i^{th} error signal for $i = 1, 2, \ldots, N$.

As an example of Eq. (3-2) consider a dynamic process characterized by the parameters $a_1, a_2, \ldots, a_n$ with $\alpha_1, \alpha_2, \ldots, \alpha_N$ representing the desired values of these parameters, respectively. If F is chosen as a quadratic function for each parameter, Eq. (3-2) becomes

$$I.P. = A_1(\alpha_1 - a_1)^2 + A_2(\alpha_2 - a_2)^2 + \ldots. + A_N(\alpha_N - a_N)^2 \qquad (3\text{-}3)$$

where the A_i are arbitrary weighting factors of each aspect of system performance.

In applying the concepts of dynamic programming to the optimization of control processes, Bellman[20] has postulated a rather broad class of I.P.'s in terms of cost functions. Bellman's development will be sketched here to add insight into the discussion to follow. It will become apparent to the reader that I.P.'s presently being used in control theory are particular cases of Bellman's formulation.

Consider a dynamic process shown in Fig. 3-1 and let the state of the process be characterized by a vector $\bar{c}(t)$ and let $\bar{m}(t)$ be the control or input vector. Further, let $\bar{c}_o(t)$ represent the desired state of the process, $G\left[\bar{c}_o(t) - \bar{c}(t)\right]$ be a function measuring the cost of deviation of $\bar{c}(t)$ from $\bar{c}_o(t)$, and $H\left[\bar{m}(t)\right]$ be a function measuring the cost of control. Then the total cost function or I.P., denoted $J\left[\bar{c}(t), \bar{m}(t)\right]$, becomes

$$J\left[\bar{c}(t), \bar{m}(t)\right] = G\left[\bar{c}_o(t) - \bar{c}(t)\right] + H\left[\bar{m}(t)\right] . \qquad (3\text{-}4)$$

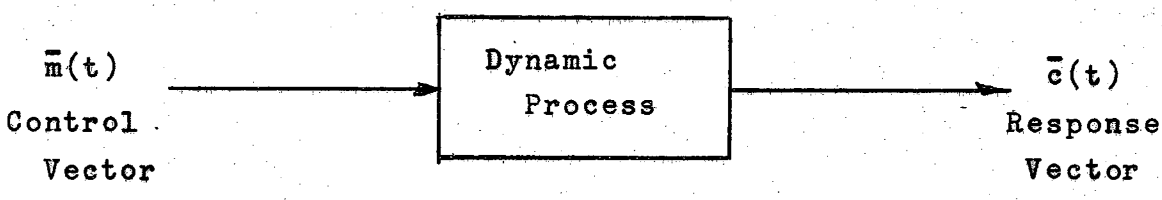

Fig. 3-1

Multi-Dimensional Dynamic Process with Control and Response Vectors

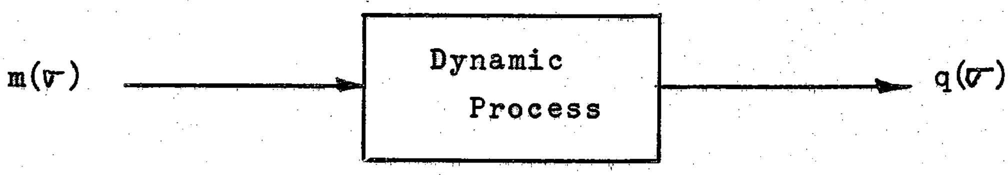

Fig. 3-2

One-Dimensional Dynamic Process with Single Input and Single Output

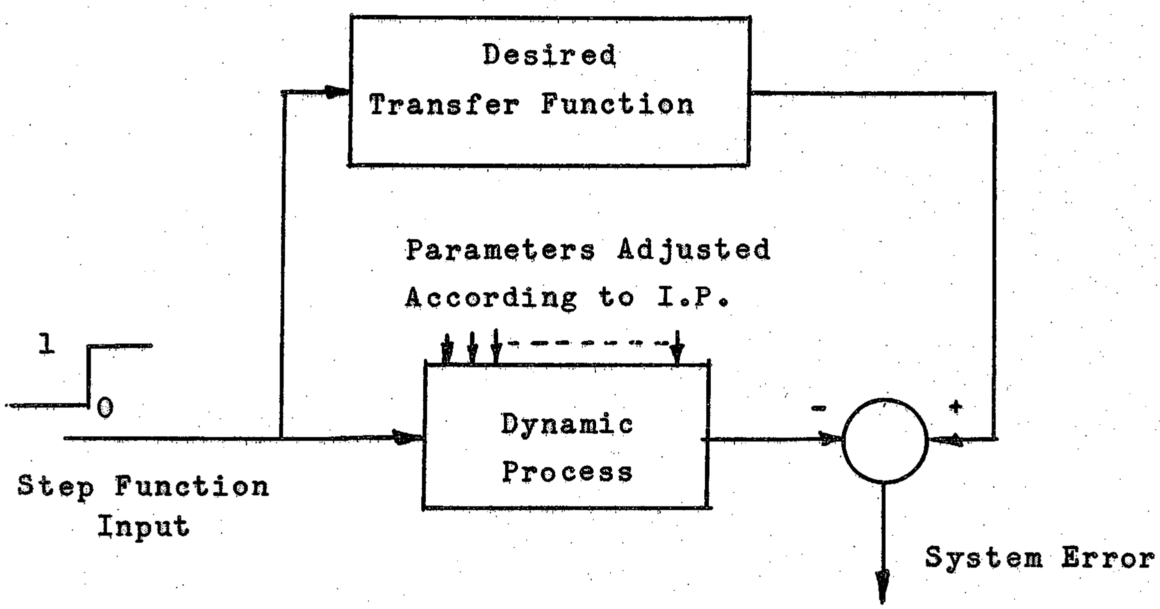

Fig. 3-3

Scheme for Obtaining System Error in Terms of Step Response

Observe that the total cost function is compounded into two parts; the first is actually a measure of system error as discussed earlier in this section, and the second is a measure of the amount of control effort to be exerted in driving the system from its present state to the desired state. Depending upon the classes of functions chosen for G and H, the optimization of Eq. (3-4) represents a variational problem of greater or lesser difficulty.

For application of Bellman's work to adaptive controls Merriam[21] has specialized Eq. (3-4) to a one-dimensional I.P. involving integrals of arbitrary functions of two system errors over a finite interval of time. Consider the dynamic process of Fig. 3-2 having a single output $q(\sigma)$ and a single input $m(\sigma)$. Let t equal the present time, and consider the interval $t \leq \sigma \leq t + \tau$ where τ is some constant. Further, let $Q(\sigma)$ and $M(\sigma)$ represent the best available estimates of the desired output and the desired input, respectively, over the specified interval. Using the above definitions, Merriam specifies the I.P.,

$$e = \int_{t}^{t+\tau} \left\{ \lambda(\sigma) f_q \left[Q(\sigma) - q(\sigma) \right] + f_m \left[M(\sigma) - m(\sigma) \right] \right\} d\sigma \qquad (3\text{-}5)$$

where

$\lambda(\sigma)$ = arbitrary weighting factor and $f_q(x)$ and $f_m(x)$ are strictly convex functions.[22] Note that Merriam does not consider the cost of control resources as does Bellman, but rather incorporates it into a second error term as indicated by the second term in the integrand of Eq. (3-5). Finally he specializes Eq. (3-5) by choosing $f_q(x)$ and $f_m(x)$ to be quadratic functions.

The I.P. in which the functional operator F of Eq. (3-1) has been chosen to be the integral over all time of the square of system error (abbreviated ISE) has been widely used as a means of defining system characteristics for deterministic input signals.[23] This I.P. is represented symbolically by

$$\text{I.P.} = \int_{-\infty}^{\infty} e^2(t)\, dt \qquad (3\text{-}6)$$

For stochastic inputs this I.P. is termed the mean-square-error[24] I.P., and is given by

$$\text{I.P.} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} e^2(t)\, dt \qquad (3\text{-}7)$$

More recently[25] an arbitrary weighting factor has been added to the integrands of Eqs. (3-6) and (3-7) to give

$$\text{I.P.} = \int_{-\infty}^{\infty} \lambda(t)\, e^2(t)\, dt \qquad (3\text{-}8)$$

and

$$\text{I.P.} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \lambda(t)\, e^2(t)\, dt \qquad (3\text{-}9)$$

respectively, where $\lambda(t)$ is the arbitrary weighting factor and has been introduced to allow unequal weighting of response errors as determined from engineering considerations.

Typical choices for the weighting factor $\lambda(t)$ have included powers of t and simple exponentials.

Other integral forms of I.P.'s which have been treated in the literature[26,27,28] are

$$I.P. = \int_0^\infty t^n \, e(t) \, dt \qquad (3\text{-}10)$$

$$I.P. = \int_0^\infty t^n \, \left| e(t) \right| \, dt \qquad (3\text{-}11)$$

Further discussion of results obtained using the above I.P.'s will be given in Section 3.3.

The use of the step function as a test signal for evaluating electronic amplifier performance in the 1940's carried over into the area of servomechanisms.[29] Throughout the 1950's the use of step response as a means of defining system characteristics came to be rather widespread primarily because it was experimentally tractable as well as analytically easy to treat. As a result many I.P.'s are defined directly in terms of step response. For linear, time-invariant, lumped parameter dynamic processes, the process transfer function is defined completely by the process impulse response, step response, or frequency response, each being directly obtainable from the other. In terms of step response the system error is obtained by utilizing a model to specify the desired transfer function, applying a unit step function to both the model and the actual dynamic process, and comparing the two output variables. Often the model is only implied by demanding ideally, perfect performance. In this case, the transfer function of the model is unity. The scheme is depicted in Fig. 3-3. The system error signal obtained is then substituted into the I.P. chosen and the I.P. minimized with respect to the system parameter adjustments which are available.

With respect to step function response, it may be desirable to control rise time, peak overshoot, rate of decay of oscillations, often termed relative damping factor, steady-state error, time for error to become less than some particular magnitude, which is sometimes called the solution time criterion or other aspects of system performance. For example, Schiewe[30] has considered a second order system in which the natural frequency, the peak overshoot, and the steady-state error to a step function input are all controlled to equal a set of predetermined values. For the second order case, the above set of parameters define the dynamic response of the system precisely. As systems of higher order are considered, however, it becomes necessary to impose more constraints upon the system to define its dynamic response adequately. The instrumentation is the same for higher order systems as it is for the second order system. If it is absolutely necessary that rise time, peak overshoot, and perhaps other aspects of performance be controlled, this approach offers one possible way to achieve this control.

3.3 Discussion of Some Results and Examples

If, in a particular application, it is required that a system have a step response which displays certain general characteristics, but which need not have a definite rise time or peak overshoot, it is possible to use an I.P. having the simple form

$$\text{I.P.} = \int_0^\infty t^n e^2(t)\, dt \qquad (3\text{-}12)$$

The value of n chosen (usually an integer) will determine, to some extent, the type of dynamic response to be expected. The amount of control of dynamic response available will depend on the number of parameters available for adjustment and the range over which they may be adjusted. The I.P. of Eq. (3-12) is applicable only to those dynamic processes whose error for a step input goes to zero at least at the rate $t^{-\sqrt{n+1+\Delta}}$ as $t \to \infty$ where Δ is arbitrarily small.

When it is desired to instrument the evaluation of this integral, it will be necessary to use some value of time T as the upper limit of integration. This upper limit must be chosen such that the essential transients due to the application of the step function have subsided. Otherwise, the value of the approximation of the integral to the exact value using the upper limit of infinity will be poor.

A large number of I.P.'s have been, and are currently being, investigated by a research group in the School of Electrical Engineering at Purdue University under Air Force sponsorship.[28] While this work is not directly concerned with the application of I.P.'s to adaptive control systems, it will be worthwhile to review some of the results because of their possible applicability for the evaluation of adaptive controls.

The Impulse Response Area Ratio (IRAR), which is the ratio of positive area to negative area in the impulse response, was not found to be of general use. It is useful for second order systems, where it is directly related to the damping ratio, but has limited application for higher order systems since it does not necessarily yield direct information about damping in these cases. Slight modification of this I.P. may be of value, however, even in higher order systems. For example, if the area of the impulse response to the first zero is compared with the total area, this gives the peak over-shoot of the system in response to a step input. Also, it is a null type I.P., or, at least, can readily be converted to yield a null at the desired value. This is an advantage since the optimum value is then known exactly. This type of I.P. shows promise for adaptive systems of higher order, although it is only of limited use as a general I.P.

The Logarithmic Decrement was found to be of no general use. It has significance only for second order systems.

The Control Area, given by Eq. (3-10) for $n = 0$, was found to be of no general use. It has significance only for second order systems.

Time Weighted Control Area, given by Eq. (3-10) for $n = 1, 2, \ldots$ was found to be of no general use.

The Integral of the Absolute Value of Error (IAE) defined by Eq. (3-11) was found to be of use for second order systems, but it has inadequate selectivity to be of general use for higher order systems.

The Integral of Squared Error (ISE) defined by Eq. (3-6) for $n = 0$ was found to be of particular interest primarily because it is mathematically convenient to apply. It has inadequate selectivity to be of general use for higher order systems.

The RMS Error was not recommended as a general figure of merit (I.P.) because it gives rise to lightly damped systems. It has often been used for mathematical convenience.

Solution Time, which is defined as the time for the step response error magnitude to drop below a particular level, was not generally recommended because it gives rise to higher order systems which are underdamped. A relatively small amplitude oscillation may persist for a long period of time.

The Integral of Time Multiplied By the Absolute Value of Error, (ITAE), defined by Eq. (3-11) for $n = 1$, was recommended as a general I.P. because it yields higher order systems with reasonable response characteristics, such as relatively small overshoot and a comparatively high degree of damping.

The results are not yet complete for a number of other I.P.'s, among them, Integral of Time Multiplied By Squared Error (ITSE) defined by Eq. (3-12) for n = 1, Integral of Squared Time Multiplied By Squared Error, (ISTSE), defined by Eq. (3-12) for n = 2, and Integral of Squared Time Times Absolute Error, (ISTAE), defined by Eq. (3-11) for n = 2. Each of these I.P.'s seem to hold some promise, and further conclusions will be given later.

It should be pointed out that some of the I.P.'s which are not generally recommended may be useful in a particular application. For example, IRAR has been used in an adaptive control system proposed and built by Aeronutronic.(31)

Graham and Lathrop(26) have carried out extensive work on the type of step response obtained using various I.P.'s for systems having transfer functions of the form

$$\frac{C(s)}{R(s)} = \frac{1}{s^m + a_{m-1} s^{m-1} + \ldots + a_1 s + 1} \qquad (3\text{-}13)$$

where C(s) is the Laplace transform of the output and R(s) the Laplace transform of the input. I.P.'s which they considered include Eqs. (3-10) and (3-11) for the lower values of n, e.g., n = 0, 1, 2, 3. Examination of Eq. (3-13) reveals that their results are restricted to systems having only poles in the transfer function. That is, all the systems considered have a transfer function whose numerator is unity. Also in specifying the values of the coefficients a_{m-1}, a_{m-2}, ..., a_1 which will give optimum step response with respect to the particular I.P. used, Graham and Lathrop do not consider the cross-coupling between these coefficients which exists for almost all physical systems. The values of the various system parameters which can be adjusted enter into each of the coefficients of Eq. (3-13). As a result it may be necessary to adjust many or all of these parameters to change one coefficient. In practice then, the adjustment of all of the coefficients to their optimum values as dictated by the minimization of the I.P. may prove impossible thereby requiring a compromise choice of the coefficients.

3.4 Some Limitations of Indices of Performance

The formulation of what might be termed the ideal I.P. has not yet been done. By the ideal I.P. we imply here one which can be applied to all control problems and which will lead to over-all system operation which is satisfactory to the user. In other words, the inexperienced systems designer should be able to apply this ideal I.P. to any problem with confidence that it will yield a reasonable answer. The use of any particular I.P. must be restricted to those classes of problems in which the I.P. has been known to give reasonable results for problems belonging to that class. Unfortunately, there does not exist any analytical means by which an I.P. may be selected for a given problem and be guaranteed to give meaningful results when interpreted physically in terms of actual system response. The difficulty arises because optimization of the system is executed with respect to the I.P. itself while the actual response is evaluated afterward in terms of other measures of performance which have been analytically impossible to include in the I.P. For example, the use of ISE (Eq. 3-6) to optimize the step response of third and higher order systems has led to "optimum" values of system parameters which produce a system whose step response has proven too oscillatory for most applications. In this example while the parameter values as obtained from optimizing the I.P. will render the ISE a minimum for a step input, they also produce a system whose step response is unsatisfactory with respect to damping. Damping, of course, is a very important measure of the quality of the step response of most systems, but is, unfortunately, impossible to express explicitly in an I.P. for higher order systems. Hence, if one is interested in optimizing the step response of such systems with respect to measures like damping, rise time, and steady-state error, the application of ISE is not to be recommended.

The ability to chose an I.P. which can be handled analytically and still satisfy all system specifications which cannot be included in the I.P. explicitly is at present a matter of experience. An I.P. which yields good results for some systems may lead to poor response for others, and the system's designer must be aware of such contingencies. In a sense then, analytical design theory for automatic control systems is not completely analytical, but requires a subjective analysis of the original problem specifications in order to proceed analytically.

I.P.'s which do not give unique results, that is, unique values for parameter adjustments, when optimized are not of general use. Such situations do not normally arise when a given I.P. is minimized, but may arise when it is specified that the I.P. be maintained at or below some fixed value. For example, consider the system having transfer function

$$\frac{C(s)}{R(s)} = \frac{1}{s^2 + 2\zeta_o s + 1} \qquad (3-14)$$

where ζ_o is the parameter to be adjusted to constrain the value of

$$\text{I.P.} = \int_0^{\infty} e^2(t)\,dt \tag{3-15}$$

to be between one and two for a step input. For this example, Graham and Lathrop[26] have plotted the curve of I.P. vs. ζ_o which has been reproduced in Fig. 3-4. Clearly, constraining the I.P. between the values of one and two leads to two values of ζ_o. However, the dynamic responses of the two systems will differ considerably, one being more oscillatory than the other. Observe that had the specification been to determine the ζ_o which minimized the I.P., then the solution would have been unique.

The selectivity of an I.P. is a function of its ability to indicate small changes in system parameters or system dynamic performance. As an example, the ability of various I.P.'s to indicate changes in the damping factor ζ_o of a second-order dynamic process might be considered. If small changes in ζ_o from the optimum are reflected by large changes in the I.P., the I.P. is considered to be selective. The more selective an I.P. is, the easier it will be to use it as a design tool, since the optimum parameter values will be more sharply defined.

In summary three limitations of I.P.'s have been indicated. The first is the usefulness, or equivalently, the applicability, of a given I.P. to a particular problem. It has been argued that the usefulness of an I.P. is some function of the original problem specifications and the choice of an I.P. for that

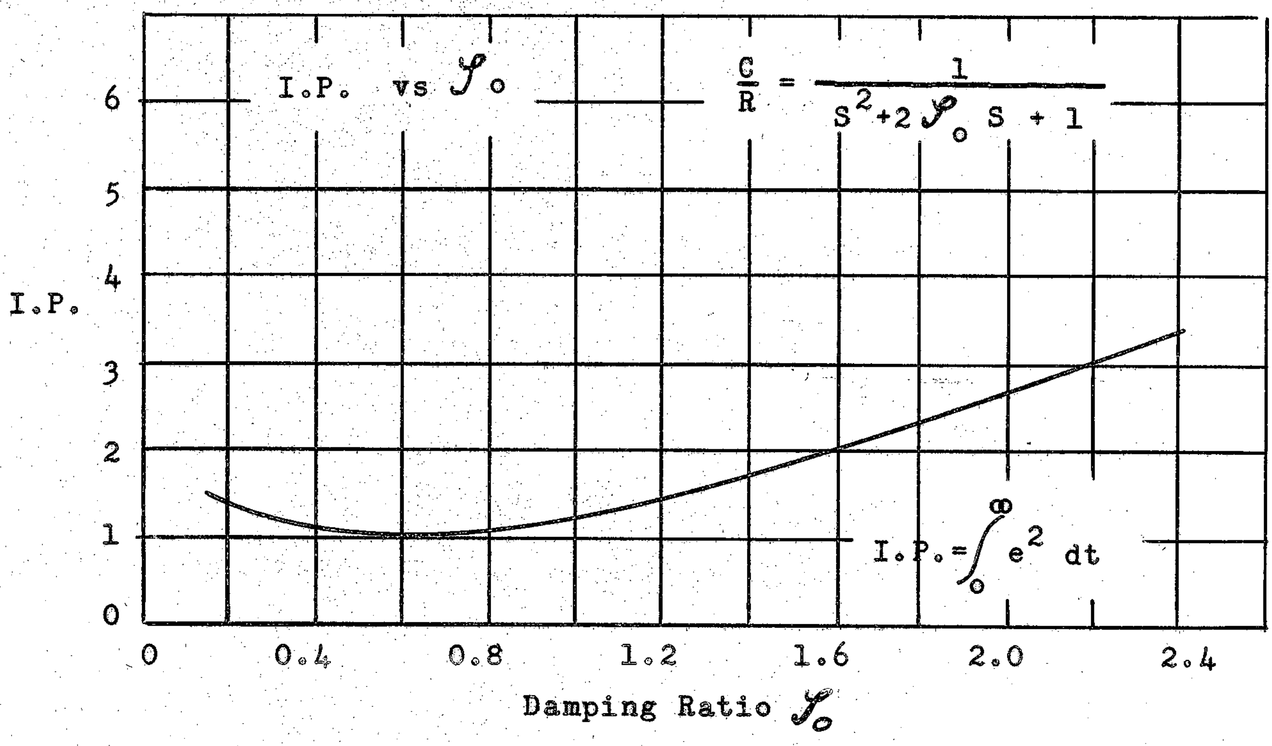

Fig. 3-4

I.P. vs. $\mathscr{S}_o$ for a Normalized Second-Order System (Reproduced from Graham and Lathrop[26]) problem is largely a matter of engineering judgment based on previous experience. Secondly, I.P.'s which do not give unique results are to be avoided unless all the results but one can be discounted by further analysis. Finally, the problem of selectivity is paramount if I.P.'s are to be used in adaptive control systems. If an I.P. is not selective to parameter changes about the optimum, its use in the adaptive loop will destroy the purpose of the adaptive loop itself.

CHAPTER IV

THE MODIFICATION PROBLEM

4.0 Abstract

The modification problem deals with methods and techniques for physically bringing the dynamic process to the optimum or desired state. This adaptation is achieved by performing linear, nonlinear and/or time-varying operations on the control signal and is termed control signal modification. Control signal modification results from system parameter adjustments or from the synthesis of completely new control signals.

Adaptation specifications are given in terms of decision requirements which give the types of adaptation to be performed and in terms of actuation requirements which state the types of adjustments to be made to realize adaptation.

This chapter includes a review of a number of recent papers representative of the state of the adaptive control science. In most cases these papers are extensions and generalizations of the earlier work completed in this area. The last section compares the two types of control signal modification and points out the economic and spatial requirements of each.

4.1 Introduction

After the identification problem and the decision problem have been solved, the actual adjustment or adaptation of the dynamic process to be controlled must be executed. The modification problem deals with methods and techniques for physically bringing the complete system to the optimum or desired state. Modification is based on a knowledge of the present state of the system as given by the identification process and on a set of predetermined indices of performance as derived from the decision process. The latter process will implicitly include a form of the desired dynamic response, usually a mathematical model used as a standard with which the actual dynamic process is compared.

In general, the modification process may be viewed as computer control of a dynamic process as shown in Fig. 4-1. The operations of the computer may range from simple arithmetic operations for the computation of indices of performance to adjustment of system parameters and then to the generation of signals used to actuate the dynamic process under control. Thus, if the computer controller of Fig. 4-1 is to be capable of performing modification over a wide range of changes in process dynamics and process signals and is, in addition, to be capable of adapting a chemical process as well as a space probe, the use of a computer the size of an IBM 704 might be required. Hence, Fig. 4-1 gives a conceptual scheme for the formulation of the modification problem, but is itself far

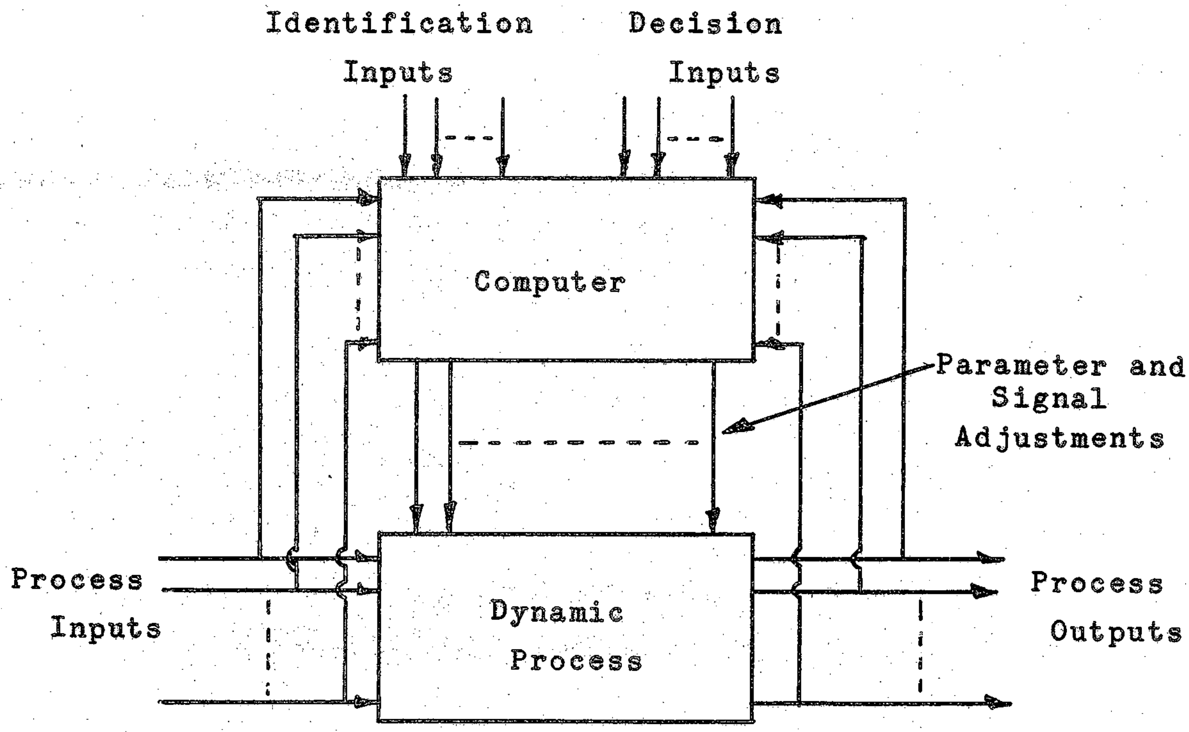

Fig. 4-1

Computer Control of a Dynamic Process

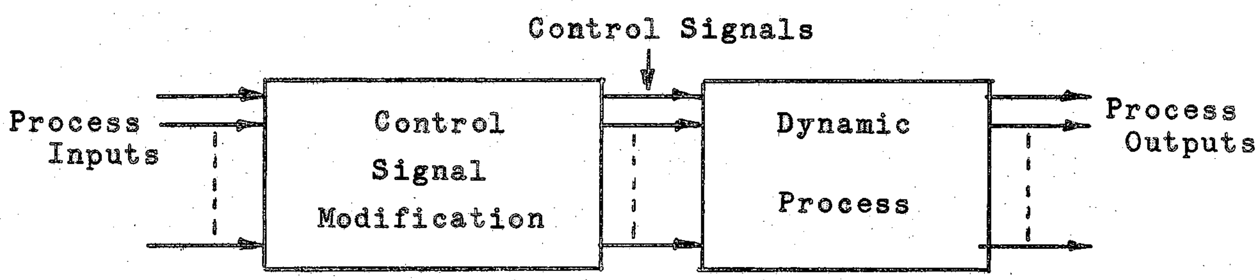

Fig. 4-2

Control Signal Modification

too general to be of any practical significance for the solution of the problem.

The approach to the solution of the modification problem presently taken by researchers working in this area is the one which employs control signal modification as depicted in Fig. 4-2. Control signal modification is defined as the application of linear, nonlinear, and/or time- varying operations to the actual system input to derive a control signal which actuates the dynamic process being controlled. This approach may be subdivided into two areas of research which have been treated in the literature but have not been distinctly defined previous to this report. These two areas are termed parameter adjustment and control signal synthesis.

Parameter Adjustment

This approach performs modification by the adjustment of the parameters of the dynamic process or a compensation network to satisfy the indices of performance as specified by the decision process. (See Fig. 4-3) Since the control requirements vary in time due to changes in process dynamics and process signals, the compensation network must have time-varying coefficients. This case is treated in detail in Section 4.3 where a literature review of the present status of the approach is given.

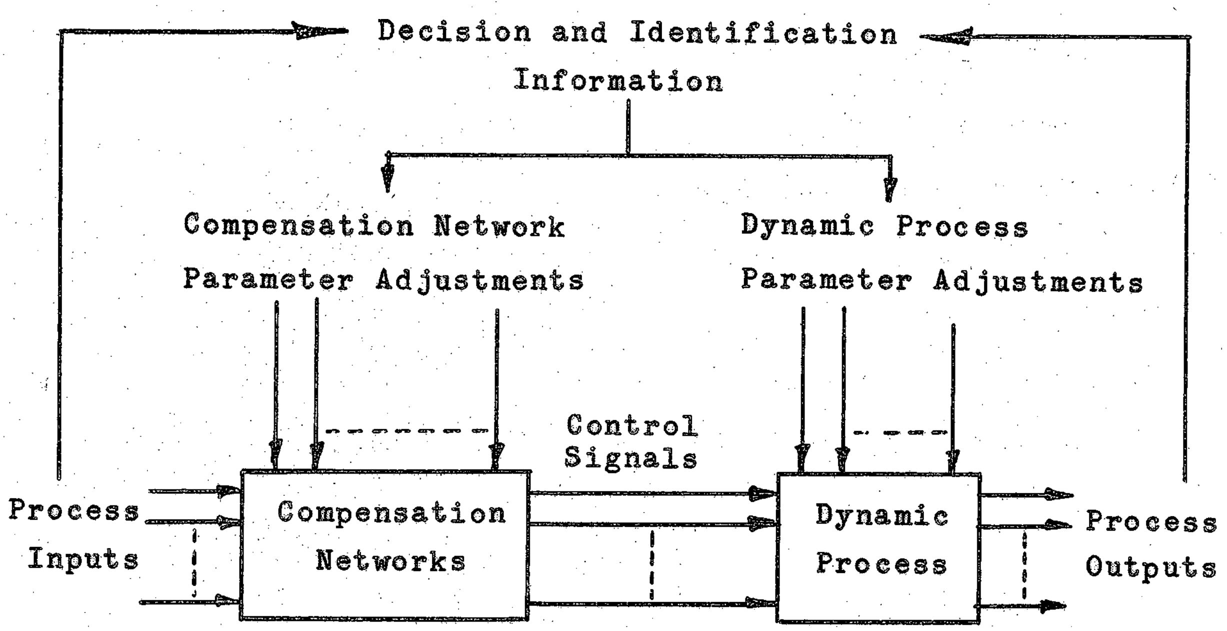

Fig. 4-3

Parameter Adjustment

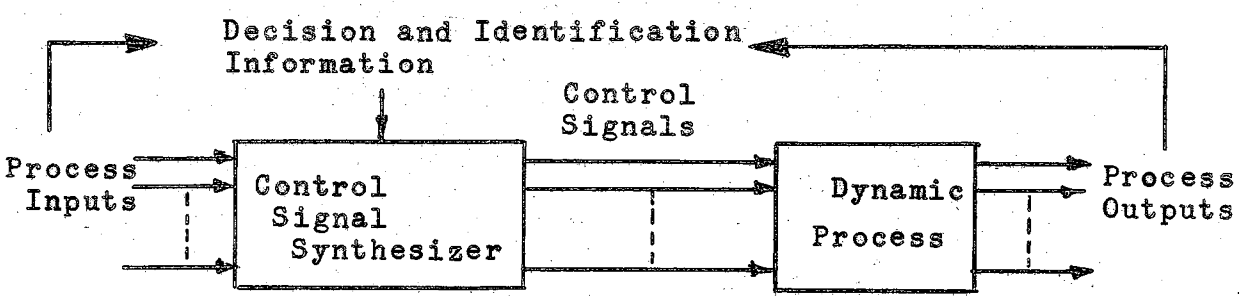

Fig. 4-4

Control Signal Synthesis

Control Signal Synthesis

Rather than perform linear, nonlinear, and/or time-varying operations on the actual system input, an alternate approach to adapting the system by control signal modification utilizes the information derived from the identification and decision processes to synthesize a new control signal which is used to actuate the dynamic process. This scheme is shown in block diagram form in Fig. 4-4. It is to be expected that the "signal synthesizer" in the system will be comprised of linear, nonlinear, and time-varying elements which may be digital and/or analog devices. This case is also treated in Section 4.3 where a literature review is included.

4.2 Adaptability Requirements

Before embarking upon a detailed treatment of the two means for achieving control signal modification, it will be worth while to review some of the underlying concepts and ideas from which the modification problem arises. Such a review will also aid the reader in understanding the motivation for the various approaches to the problem taken by different authors. These ideas and concepts are embodied in a set of adaptability requirements or specifications which plays an integral role in the formulation of the modification problem. Adaptability requirements fall into two categories: the first dealing with the types of system changes to which it is desired to adapt, and the second delineating the types of adjustments to be made to achieve adaptation. These requirements will be termed decision requirements and actuation requirements, respectively.

Decision Requirements

The physical realization of adaptation cannot be initiated until a decision has been made as to the types of changes to which the system is to adapt. These types of changes are categorized as:

1. Changes in process dynamics.
2. Changes in the statistics or deterministic character of the signals present in the dynamic process.

3. Changes in the type of internal and external disturbances present in the dynamic process.

More generally, these changes may be viewed as changes in the system's environment.

In any practical application the system may be called upon to adapt to any one or any combination of the types of changes listed above. Once this specification has been made, the choice of just how to achieve adaptation to the different changes must be selected. This leads naturally to the specification of actuation requirements.

Actuation Requirements

The diversity of applications of adaptive control systems mandates the subdivision of actuation requirements into the following three classes which represent the current approaches to adjustments for realizing adaptation. These cases are:

1. Adjustment to optimum operating points.
2. Adjustment of system parameters to achieve a desired dynamic response according to predetermined indices of performance.
3. Adjustment of system signals to cause a desired response according to predetermined indices of performance.

As in the case of decision requirements, any practical adaptive control configuration may be called upon to satisfy any one or any combination of the above actuation requirements.

Case 1 requires knowledge of the desired operating points. For example, consider control of a chemical process. Once the transients have subsided, the requirements of control are essentially those of maintaining steady-state operating points, e.g., temperatures, pressures, and flow rates, to achieve the desired quantity and quality of the process products. Adaptive control in this case would be concerned with achieving and maintaining the desired steady-state operation in a minimum time with a minimum of loss in output products.

Case 2 is actually an extension of Case 1 and is concerned with control in order to achieve a desired dynamic response rather than a steady-state behavior. An example in point here would be the adjustment of the parameters of a radar detection system to maximize the signal-to-noise ratio in the output and at the same time minimize integral-square error to a ramp input in the presence of sporadic atmospheric disturbances and changes in the types of objects being tracked. That is, the system should possess enough adaptability to track a high altitude reconnaissance plane as well as it does a space vehicle despite the presence of noise inputs whose statistics are time-varying. Clearly, this case deals with adjustment of system parameters themselves rather than system signals, although it is obvious that adjustment of system parameters will alter the behavior of signals. The use of a mathematical model specifying the optimum adjustment of system parameters to achieve the desired dynamic response is implicit in Case 2.

Case 3 is in contrast to Case 2 because it deals with altering the signals present directly, rather than attempting to adjust parameters of the system. In many applications it may be impossible to alter the character of the dynamic process or of a compensation scheme in order to achieve optimum operation. Under such circumstances, adaptation can be realized by developing new signals which may be used as corrections to those already present or as alternate sources of excitation for the dynamic process. Since adjustment is made so that the actual output of the process follows a desired response, the presence of a model for the dynamic process is explicit in this case.

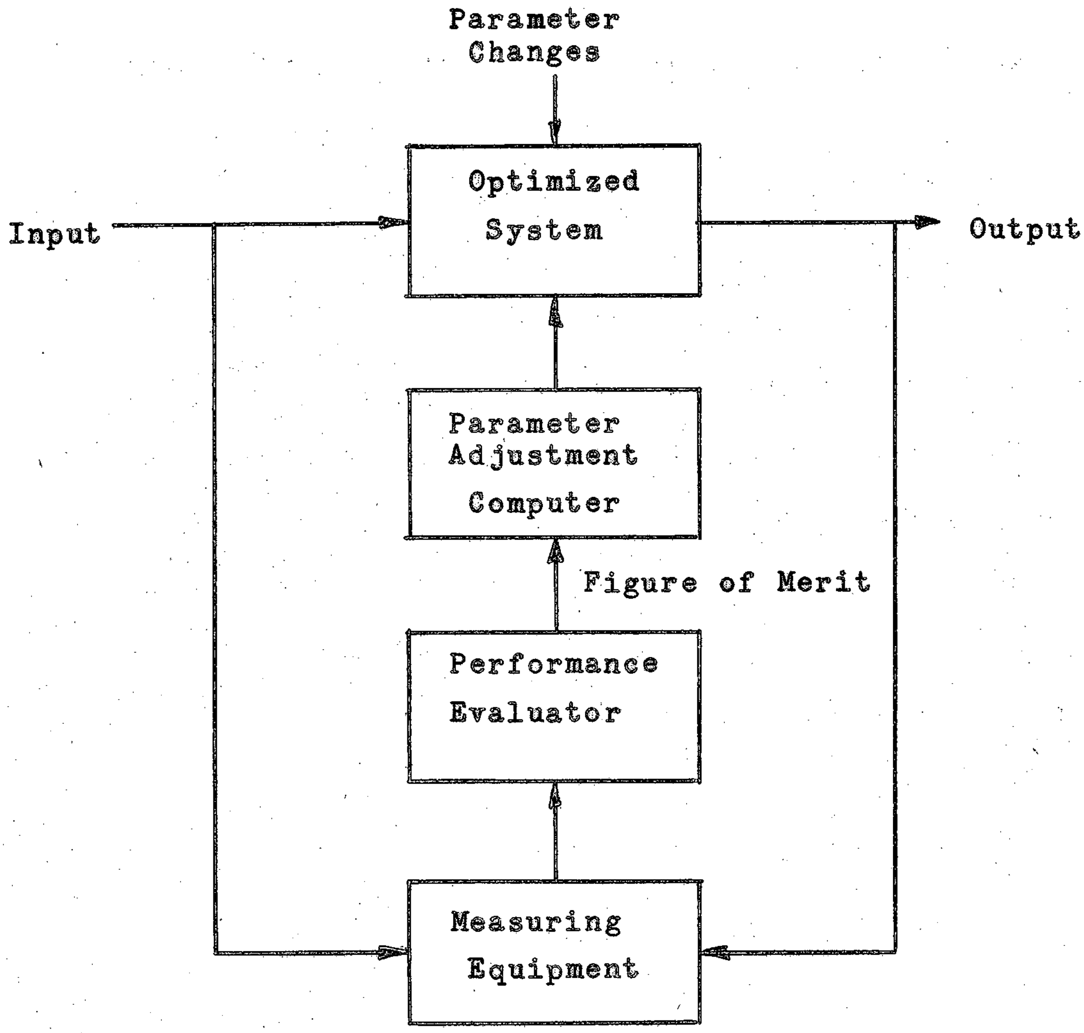

Fig. 4-5

Parameter Adjustment Adaptive Control System of Anderson, et. al.[33]

4.3 Control Signal Modification

Having defined control signal modification and reviewed the requirements which underlie adaptation, we now examine the current solutions proposed for the modification problem. The emphasis here will be on the more recent developments, more particularly, on work completed within the last two years. For a detailed bibliography of the research done in the adaptive control area the reader is referred to a paper by Stromer.[32] The presentation of the material here will parallel the subdivision of control signal modification into the two areas of research as given in Section 4.1. It will become apparent as the work is presented that almost without exception the authors rely on the use of only one index of performance as a means of evaluating system behavior. As the adaptive control science progresses, it is felt that the use of more than one index of performance will be necessary in those cases where more than one aspect of system performance is to be controlled.

Parameter Adjustment

One of the most noteworthy efforts in this area is presented in a paper by Anderson, et. al.[33] The system which is shown in Fig. 4-5 utilizes the impulse response area ratio (See Chapter III) as the index of performance. A detailed study of this method as applied to a second-order system gave very good results. The technique provides means for the system to adjust its parameters for optimum dynamic response by using a null-type index of performance.

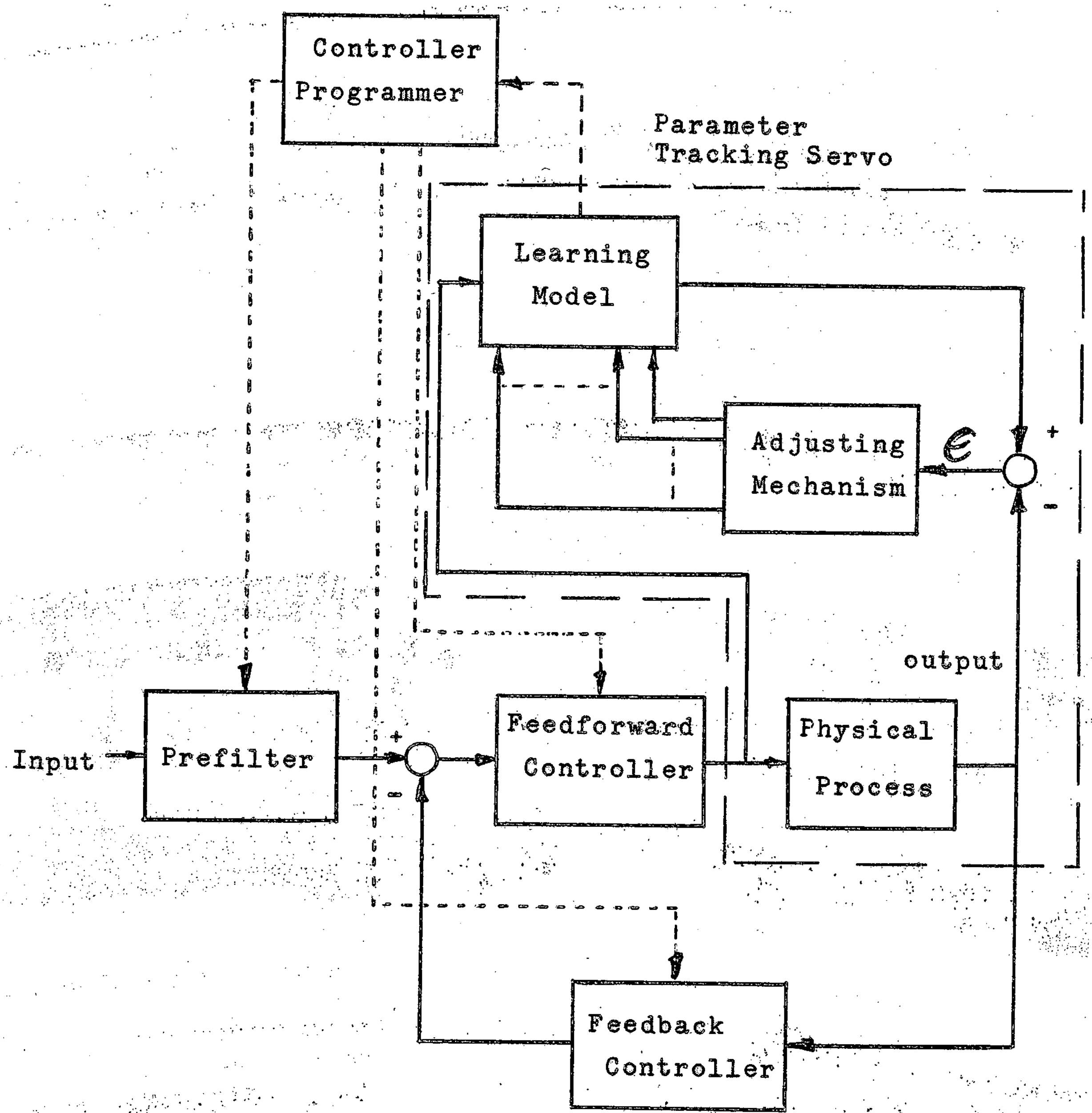

Fig. 4-6

Parameter Adjustment Adaptive Control System of Margolis and Leondes[14]

Kalman[34] optimizes the complete system by calculating the pulse transfer function and adjusting parameters in order to achieve zero error in minimum time for a step input. The pulse transfer function is used as an approximation to the system impulse response. The system is restricted severely because only step inputs can be handled.

Margolis and Leondes[14] employ a "learning model" in a parameter tracking adaptive control configuration. The scheme is shown in the block diagram of Fig. 4-6. The same signal is applied to both the learning model and the physical process whose outputs are compared to obtain an error signal. A function of this error is used to adjust the parameters of the learning model. The purpose of the adaptive loop is to track the physical process parameters continuously as they change in order to supply information to the controller programmer which then adjusts the feedforward and the feedback controllers, and the prefilter to achieve a prescribed dynamic response. The method of steepest descent[35] is used to adjust parameters. The paper treats only the first order physical processes, but work to extend the method to higher order processes is under consideration.

The problem of applying techniques from dynamic programming[20] to realize parameter adjustment is considered by Bellman and Kalaba.[36] The authors illustrate the concepts by considering a process which is governed by the inhomogeneous Van der Pol equation

$$\ddot{x} + \mu (x^2 - 1) \dot{x} + x = r(t), \; 0 \leq t \leq T. \qquad (4\text{-}1)$$

The adaptive loop is called upon to maintain the process near the state $x = 0$, $\dot{x} = 0$ by adjusting μ. The function r(t) is a random function whose statistical properties are not completely known at the outset. As the process unfolds, the adaptive loop is able to obtain more information about the statistical properties of r(t) and therefore improve the adjustment of μ to maintain the desired state of the process. The particular example considered is basic in describing relaxation oscillations in vacuum tube oscillators and in multivibrators.

Chang(37) has utilized Z-transform methods to achieve parameter adjustment. The problem of maintaining a parameter at a prescribed value or at some unknown extremal value is considered. The index of performance used for maintaining the parameter at a prescribed value is rms error; that for extremal seeking systems is least reduction in the parameter. The author considers the problems associated with finite measuring time, probable error of measurement, and effects of large changes or disturbances in the parameter being controlled.

Control Signal Synthesis

As pointed out earlier, it may be impossible to perform modification by parameter adjustment in a large number of applications. This situation will arise in those cases where system parameters must be measured and controlled indirectly because the dynamic process has no physical adjustments available. It will also arise in those cases where the adaptation requirements are so severe that adjustment of the parameters of a compensation scheme is also inadequate to account for all contingencies. The application of control signal synthesis has provided a powerful means of performing modification under the above conditions.

The philosophy of control signal synthesis as presented in Section 4.1 is the foundation upon which the research efforts reviewed below are predicated. A number of the plans proposed do not consider constraints on the control variable to prevent saturation. As a result, their applications are limited.

However, in contrast to the work done in the parameter adjustment area, the research done in the control signal synthesis area has been concerned with the overall system response as well as the response of the adaptive portion of the system. Little research effort has been devoted to the response of the overall system in the former area.

Braun's [8] method makes use of the Maclaurin series expansion for the dynamic process impulse response, the process forcing function, and the process output. With this knowledge, the adaptive loop proceeds to synthesize a new signal which when added to the actual process forcing function will constitute the necessary correction to force the process output to follow the desired process output exactly. The corrective signal is the form of a sum of a finite number of singularity functions and includes an impulse which, of course, if applied will violate the linearity of the dynamic process. The procedure is repeated every T seconds where T is the amount of time necessary to determine whether or not a change has occurred in the process impulse response. A recursion relation is developed from which the coefficients of the terms in the corrective signal are computed by digital means. The effects of computation time and computation errors are not considered. Moreover, Braun does not state the source from which the desired process reponse is obtained.

An extension of the concepts of dynamic programming has been made by Merriam[38] to obtain an optimum adaptive control configuration which employs time-varying gains in a feed-forward and feedback scheme to achieve modification. A modified least squares index of performance is postulated, and dynamic programming procedures are applied to determine what the dynamic process input must be in order that the actual process output will approximate the desired process output in the least squares sense. It is found that the optimum process input can be derived from the desired process response and the actual process response through time-varying gains. Unfortunately, these time-varying gains must be obtained by solving a set of simultaneous nonlinear differential equations. In his thesis[39] Merriam considers third and higher order dynamic processes to which the above scheme can be applied.

However, only for first-order processes does he consider constraints on the process input.

Freimer[40] has also applied dynamic programming notions to a class of control processes which obey a differential equation of the form

$$\dot{x}(t) + A(t)\, x(t) = y(t) \tag{4-2}$$

where x(t) is a real s-dimensional control vector and A(t) is a known s x s matrix function of time. The problem is one of choosing y(t) to minimize an error functional. Stochastic and deterministic control situations are treated and then specialized for a quadratic error functional to illustrate the theory.

The papers reviewed in this section were selected with the intent of summarizing the present status of the modification problem. They represent the more recent results and in most cases are extensions and generalizations of the earlier efforts in this area.

It is interesting to observe that no work has been done on the stability analysis of any of the proposed modification schemes. Clearly, the question of stability will, of necessity, arise in the evaluation of any closed-loop control configuration. Since the adaptive loops employed in these systems are at best nonlinear, stability analysis will not be simple. Indeed, many of the configurations employ computation as a control element of the adaptive loop. Before any stability analysis can be effected, a suitable characterization of that portion of the adaptive loop must be developed.

In addition, more analytical and experimental work is needed to compare the overall system response or behavior with that of conventional control systems.

4.4 A Comparison of Parameter Adjustment and Control Signal Synthesis

In summarizing the results of this chapter, it will be worth while to give a brief comparison of the two current approaches to the solution of the modification problem. In principle the two are equivalent since they are both particular cases of the general notion of control signal modification. However, there do exist differences which make one approach more facile than the other in a given application. A particular class of applications in which control signal synthesis is to be preferred to parameter adjustment has already been indicated in Section 4.3. Even where parameter adjustments are available, they may not provide the flexibility necessary to obtain the type of dynamic or steady-state behavior required by the indices of performance. The type of control signal modification to be used in any engineering application will depend in part on the type of adaptation to be performed. In addition, it will depend on economic and spatial considerations as discussed in the following three paragraphs. Under severe adaptation requirements, a combination of the two approaches may prove to be the only solution.

Present day adaptive control technology places heavy demands on digital and/or analog computation. The questions of economics and space requirements are explicit in the choice of the computational facility which forms the nucleus of the adaptive loop. One is faced with the choice between a large digital facility capable of accurate, high-speed computations and an analog facility which would employ nonlinear and time-varying operations with consequential losses in accuracy. In addition to representing a large financial investment, the digital facility will generally be quite heavy and require considerable space. Hence, unless a small, special purpose, solid-state digital computer capable of performing the required operations can be built, adaptive control systems employing digital computation are impractical for airborne applications. On the other hand, if moderate losses in computing accuracy can be tolerated, analog facilities can be built which would keep weight and size at a minimum. In some instances a compromise engineering design utilizing both digital and analog devices may be possible. Clearly, weight and size are crucial factors if the system is to be airborne.

In addition, a choice between rapid and real-time computation must be made with the latter offering more simplicity of design, but again poorer accuracy than the former. Here again a compromise engineering design which sacrifices accuracy and speed for size, weight, and cost may prove necessary.

By way of a comparison, the system of Anderson, et. al.[33] proves to be far easier to instrument than Merriam's[38] configuration which will require a high-speed digital facility for the solution of the differential equations from which the time-varying gains are determined. However, the range of environmental changes over which Merriam's system will adapt is far greater than that which the system of Anderson, et. al., spans.

BIBLIOGRAPHY

(1) Aseltine, J. A., Mancini, A. R., and Sarture, C. W., "A Survey of Adaptive Control Systems," IRE Transactions on Automatic Control PGAC, Vol. AC-6, Dec., 1958.

(2) Truxal, John G., "Trends in Adaptive Control Systems," Proceedings of the National Electronics Conference, Vol. 15, 1959.

(3) Zadeh, L. A. "On the Identification Problem," IRE Transactions on Circuit Theory PGCT, Vol. CT-3, No. 4, Dec., 1956.

(4) Woodrow, R. A., "Data Fitting with Linear Transfer Functions," Journal of Electronics and Control, Vol. VI, No. 5, May, 1959.

(5) Moore, E. F., "Gendanken Experiments on Sequential Machines," Automata Studies, Princeton University Press, Princeton, N.J., 1955.

(6) Turin, G. L., "On the Estimation in the Presence of Noise of the Impulse Response of a Random Linear Filter," IRE Transactions on Information Theory - PGIT, Vol. IT-3, No. 1, March, 1957.

(7) Braun, L., Jr., "On Adaptive Control Systems," Doctoral Dissertation, Polytechnic Institute of Brooklyn, 1959.

(8) Braun, L., Jr., "On Adaptive Control Systems," IRE Transactions on Automatic Control - PGAC, Vol. AC-4, No. 2, Nov., 1959.

(9) Levine, M. J., "Optimum Estimation of Impulse Response in the Presence of Noise," IRE Transactions on Circuit Theory - PGCT, Vol. CT-7, No. 1, March, 1960.

(10) Woodrow, R. A., "On Finding a Best Linear Approximation to System Dynamics from Short Duration Samples of Operating Data," Journal of Electronics and Control, Vol. VII, No. 2, Aug., 1959.

(11) Cooper, G. R., "An Analog Method of Measuring Impulse Response," Intracompany Memorandum, Aeroutronic Systems, Inc., August 11, 1959.

(12) Kalman, R. E., "Design of a Self-Optimizing Control System," ASME Transactions, Vol. 80, Feb., 1958.

(13) Corbin, R. M., and Mishkin, E., "On the Measurement Problem in Adaptive Systems Utilizing Analog Computer Techniques," Research Report R-699-58, PIB-627, Microwave Research Institute, Polytechnic Institute of Brooklyn.

(14) Margolis, M., and Leondes, C. T., "A Parameter Tracking Servo For Adaptive Control Systems," IRE Transactions On Automatic Control - PGAC, Vol. AC-4, No. 2, Nov., 1959.

(15) Whitaker, H. P., Yamron, J., and Kezer, A., "Design of Model-Reference Adaptive Control Systems for Aircraft," Report R-164 MIT Instrumentation Laboratory, Sept., 1958.

(16) Anderson, G. W., Buland, R. N., and Cooper, G. R., "Use of Crosscorrelation in An Adaptive Control System," Proceedings of the National Electronics Conference, Vol. 15, 1959.

(17) Anderson, G. W., Aseltine, J. A., Mancini, A. R., and Sarture, C. W., "A Self-Adjusting System for Optimum Dynamic Performance," IRE National Convention Record, Part 4, 1958.

(18) Cooper, G. R., and Gibson, J. E., "Third Monthly Progress Report," Contract AF 33(616)-6890, Project No. 8225, WADD, April, 1, 1960.

(19) Newton, G. C., Jr., Gould, L. A., and Kaiser, J. F., Analytical Design of Linear Feedback Controls, John Wiley and Sons, Inc., New York, 1957.

(20) Bellman, R., Dynamic Programming, Princeton University Press, New Jersey, 1957.

(21) Merriam, C. W., III, "Synthesis of Adaptive Controls," Sc. D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, May, 1958.

(22) Courant, R., Differential and Integral Calculus, Vol. II, Interscience Publishers, Inc., New York, 1949.

(23) Newton, G. C., Jr., Gould, L. A., and Kaiser, J. F., Analytical Design of Linear Feedback Controls, John Wiley and Sons, Inc., New York, 1957, Ch. 2.

(24) Ibid., Ch. 4 and 5.

(25) Murphy, G. J., and Bold, N. T., "Optimization Based on a Square-Error Criterion with an Arbitrary Weighting Function," IRE Transactions - PGAC, Vol. AC-5, No. 1, January, 1960.

(26) Graham, D., and Lathrop, R. C., "The Synthesis of Optimum Transient Response Criteria and Standard Forms," Transactions AIEE, Vol. 72, Pt. II, 1954.

(27) Stone, C. R., "A Mathematical Method for Optimizing the Gains of a Linear System," Minneapolis Honeywell Co., Aero Report 49174-21TRI, February, 1957.

(28) Gibson, J. E., Leedham, C. D., McVey, E. S., and Rekasius, Z. V., "Specifications and Data Presentation in Linear Control Systems," AF Technical Report No. 1, Contract AF 29(600)-1933, PRF 2030, Holloman AF Base, July, 1959.

(29) Hall, A. C., Analysis and Synthesis of Linear Servomechanisms, Technology Press, Cambridge, Massachusetts, 1943.

(30) Schiewe, A. J., "A Three Dimensional Control System," Ph.D. Thesis, School of Electrical Engineering, Purdue University, Lafayette, Indiana, June, 1960.

(31) Anderson, G. W., Buland, R. N., and Cooper, G. R., "The Aeronutronic Self-Optimizing Automatic Control System," Proceedings of the Self-Adaptive Flight Control Systems Symposium," WADC Technical Report, 59-49, March, 1959.

(32) Stromer, P. R., "Adaptive or Self-Optimizing Control Systems - A Bibliography," IRE Transactions - PGAC Vol. AC-4, No. 1, May, 1959.

(33) Anderson, G. W., et. al., "A Self-Adjusting System for Optimum Dynamic Performance," 1958 IRE National Convention Record, Part 4, pp. 182-190.

(34) Kalman, R. E., "Design of a Self-Optimizing Control System," ASME Transactions, Vol. 80, Feb., 1958, pp. 468-478.

(35) Beckenbach, E. F., Modern Mathematics for the Engineer, McGraw-Hill Book Co., Inc., New York, 1956, pp. 448-478.

(36) Bellman, R., and Kalaba, R., "On Adaptive Control Processes," IRE Transactions - PGAC, Vol. AC-4, No. 2, Nov., 1959.

(37) Chang, S. S. L., "Optimization of the Adaptive Function by the Z-Transform Method," Technical Report 400-6, New York University, New York, N.Y., June, 1959.

(38) Merriam, C. W., III, "Use of a Mathematical Error Criterion in the Design of Adaptive Control Systems," AIEE Transactions Paper No. 59-1158, August 14, 1959.

(39) Merriam, C. W., III, "Synthesis of Adaptive Controls," Sc. D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, May, 1958.

(40) Freimer, M., "A Dynamic Programming Approach to Adaptive Control Processes," IRE Transactions - PGAC, Vol. AC-4, No. 2, Nov., 1959.

E