5-1-1961

Specification and Data Presentation in Linear Control Systems-Part Two

John E. Gibson
Purdue University

Eccles S. McVey
Purdue University

C. D. Leedham
Purdue University

Rekasius Z. V.
Purdue University

Schultz D. G.
Purdue University

See next page for additional authors

Authors

John E. Gibson, Eccles S. McVey, C. D. Leedham, Rekasius Z. V., Schultz D. G., and Sridhar R.

J. E. Gibson

PURDUE UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

Specification and Data Presentation
in Linear Control Systems-Part Two

J. E. Gibson, Principal Investigator

E. S. McVey, Assistant Principal Investigator

C. D. Leedham

Z. V. Rekasius

D. G. Schultz

R. Sridhar

Control and Information Systems Laboratory

May, 1961

Lafayette, Indiana

SPECIFICATION AND DATA PRESENTATION

IN LINEAR CONTROL SYSTEMS - PART TWO

J. E. Gibson, Principal Investigator

E. S. McVey, Assistant Principal Investigator

C. D. Leedham

Z. V. Rekasius

D. G. Schultz

R. Sridhar

SCHOOL OF ELECTRICAL ENGINEERING

PURDUE UNIVERSITY

Lafayette, Indiana

May, 1961

PREFACE

This report was prepared by Purdue University, School of Electrical Engineering, Prof. J. E. Gibson acting as Principal Investigator, under USAF Contract No. AF 29(600)-1933. This contract is administered under the director of the Guidance and Control Division, Air Force Missile Development Center, Holloman Air Force Base, New Mexico, by Mr. J. H. Gengelbach, the initiator of the study.

ABSTRACT

This is the second part of a 2 volume report on the specification and data presentation in linear control systems. This volume deals with Sample Data Systems, Linear Time Varable Parameter Systems, and Performance Indices, which are respectively Chapter II, III, and IV of the volume. Since these subjects are somewhat unrelated, a separate abstract is given at the beginning of each chapter, with the exception of the introductory Chapter I. The separate chapter abstracts are repeated here for the convenience of the reader.

Abstract - Linear Sampled Data Control Systems

The specifications recommended for use with sampled data control systems are those recommended for linear, continuous systems [1]. These specifications must be supplemented, as is dictated by the requirements of a particular system, by compatibility considerations that are detailed in the following sections.

Abstract - The Specification of Linear Time Variable Parameter Systems

Linear time variable parameter (LTVP) systems are defined and subdivided into those systems with fast or slow variations and/or large or small variations. The methods of analysis of such systems are reviewed, and the following recommendations are made.

Specifications

1) Time Domain Specifications

   (a) LTVP systems with fast variation of parameters.

   Simulated unfrozen system step function responses should all lie within a prescribed envelope. Whenever possible, the actual system response should be obtained.

   (b) LTVP systems with slow variation of parameters.

Simulated or actual frozen or unfrozen system step function responses should all lie within a prescribed envelope.

2) Frequency Domain Specifications

(a) LTVP system with fast variation of parameters.

Frequency domain specifications are not recommended.

(b) LTVP system with slow variation of parameters.

The family of frequency response curves of the system frozen at different instants should all lie within a predetermined envelope.

Data Presentation

It is recommended that the region of variation of closed loop poles of the frozen system be exhibited on the complex plane. Thus, for example, if the only varying parameter is an open loop gain, then the region of variation of the closed loop poles will correspond to the root loci over the total range of variation of gain.

It is also recommended that a family of Nyquist diagrams corresponding to the system frozen at different instants be displayed in the case of system with slow variations of parameters.

Abstract - Performance Index

This study was undertaken to determine whether or not Performance Indices should be used to evaluate and specify control systems. It is recommended that they not be used at this time by the Air Force for the stated purpose.

A performance index is defined and detailed discussions are presented for the various performance indices. Analytical methods for evaluating performance indices are presented.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This is the second volume of the final report on this contract, and this volume, as indicated by its title, is a continuation of Volume I, "Specification and Data Presentation in Linear Control Systems", published in October, 1960. Volume III, entitled "Stability of Nonlinear Control Systems by the Second Method of Liapunov" is the third and last volume of the final report and is to be published along with Volume II. An interim report on specifications for nonlinear systems will be published shortly.

In this Volume II three topics of considerable importance in linear systems are discussed, and these are Sampled Data Systems, Linear Time Variable Parameter Systems, and Performance Indices.

Sampled data systems may be considered linear if amplitude quantization distortion is neglected, and this is the position taken almost universally in the analysis of such systems. Sampled data systems have received considerable attention in the technical literature in the past decade. This attention has not always been because of the practical importance of such systems, but often because of the interesting mathematics that are involved. In other words the analysis of sampled data systems has become an academic discipline much like network synthesis in character. The sampled data system that is designed to compete with a continuous system must be judged by the same performance criteria, it would seem, and this is the point of view of this volume. In some cases a sampled data subsystem is to be procured that must be compatible with an overall system. Naturally, then, this subsystem must meet compati-

bility requirements on sampling rate and so forth, over and above meeting
certain performance specifications. Compatibility requirements must al-
ways be met, of course, but this is a specialized area and outside the
bounds of this study, and for this reason that topic is not considered
here.

The analysis and specification of Linear Time Variable Parameter
(LTVP)systems is one of great interest and importance to the Air Force.
The sad state of this art, even though the almighty law of superposition
still applies, should serve as a great source of embarrassment to applied
mathematicians and engineering scientists. Apparently it is only very
recently that attempts have been made to apply modern operational tech-
niques, to this problem. The state of the art in this area is discussed
in Chapter 3 of this volume.

One of the early hopes in this research was that generalized Indices
of Performance or Figures of Merit could be developed for control systems.
It was hoped that Indices of Performance would do two things; First, per-
mit generalized design procedures based on these criteria to be worked
out for linear systems, and Second, permit the comparison of two or more
competitive systems by the Air Force, so as to aid in the objective
evaluation of competing designs. Some progress has been made in this
direction and is reported in Chapter 3. However, it does not appear at
this time that such a procedure will ever be completely successful.
This is so because the relative weighting of the various factors that go
into such an index depend not only upon the operational requirements for
the system but also upon the design philosophy and judgment of the vendor
and the buyer. This does not mean, however, that Performance Indices
will not become more widely used than they are at present as their merits

become better known. It is simply that in the opinion of the Purdue group no one criterion can ever be the universal solvent or magic wand.

The interim report on specifications for nonlinear control systems will outline an approach to the problem and will discuss the state of the art. This report will then be circulated to Air Force vendors for criticisms. This approach was first suggested by AFMDC and was followed in the first portion of this work. The reaction of the vendors was favorable and a number of changes were incorporated in the final report as a result of this feedback. It will not be possible to carry that work to its conclusion and issue a final report on that material within the confines of the present contract.

For the convenience of the reader the specifications recommended in Volume 1 of this Final Report [1] are reproduced here. These specifications fall into two groups: Frequency Domain and Time Domain.

1) The frequency domain specifications are to be measured for sinusoidal input frequencies. The recommended specifications are:

    M-Peak, M_P

    Peak Frequency, ω_p

    Bandwidth, B. W.

    Peak Output Impedance, Z_p

II) The time domain specifications are to be measured at the output terminals for step inputs. The recommended specifications are:

    Delay Time, T_D

    Rise Time, T_R

    Peak Overshoot, PO

    Settling Time, T_S

    Final Value of Error, FVE

These specifications can be expressed together with their toler-
ances in a convenient graphical form as an acceptable region in the
magnitude-time or magnitude-frequency spaces ([1], Figs. 3-5 and 4-1).
The system time and frequency responses can be constrained in regions
determined from the recommended specifications and with the required
performance of a particular system in mind.

CHAPTER 2

LINEAR SAMPLED DATA CONTROL SYSTEMS

Abstract

The specifications recommended for use with sampled data control systems are those recommended for linear, continuous systems [1]. These specifications must be supplemented, as is dictated by the requirements of a particular system, by compatibility considerations that are detailed in the following sections.

2.1 Introduction

A Sampled Data Control System is defined (Tou [2], p.5) as a system "in which the control signal in a certain portion of the system is supplied intermittently at a constant rate". Alternatively, systems of this type are defined (Truxal [3], p. 500) as "systems for which the input (or the actuating signal) is represented by samples at regular intervals of time, with the information ordinarily carried in the amplitude of the samples", or by "systems in which the data appear at one or more points as a sequence of numbers or as pulses are known as sampled data systems" (Ragazzini [4], p. 1). These definitions cover the group of systems under consideration; those due to Tou [2] and Ragazzini and Franklin [4] are the broader definitions for they include first, those systems defined by Truxal [3] that are pulse amplitude modulated and secondly, those systems that are pulse code modulated.

Amplitude modulated sampled data systems are those where the signal is represented by a train of pulses, ideally impulses (Linvill [5]), with the information contained in the magnitude of the impulse. Signals of this type are generated from continuous or analog data by means of

cyclic switches, and an analysis of the sampling operation ([3], p. 509)
indicates that the sampler is a linear device.

A pulse code modulated system is a sampled data system where the
signal values at the sampling instants are quantized and coded. The
signal information is thus transmitted in trains of pulse groups
during the sampling period, usually for processing by a digital computer.
Systems with this type of signal representation are called digital
control systems, and the sequential procedure of sampling, quantizing
and coding (usually binary coding) is called analog-to-digital conver-
sion. It is apparent that systems including analog-to-digital conversion
will usually include the inverse operation of digital-to-analog conversion.

The quantization process necessary for analog to digital conversion
is a nonlinear operation in the sense that only discrete levels of output
are possible, and consequently the principle of superposition does not
apply. The nonlinearity of the quantizer can be measured in terms of
the r.m.s. quantization error ([2], p. 87), which in turn depends on the
size of the quantization step. Conversion units that are finely
quantized, thus reducing the quantization error, can frequently be con-
sidered as linear elements. Principles for the determination of system
linearity have been outline in Chapter 2, Final Report, Volume 1 [1] and
can be applied to the over-all system in which a conversion unit is
included.

Systems that fail to meet the specification for linearity are out-
side the scope of the chapter.

Sampled Data systems with digital computers included in the loop
for compensation or other purposes may have associated with them finite

computing times. When viewed from the input-output terminals this constiutes a time delay in the loop. Whether such systems are to be considered as included for discussion in this chapter rests, once again, with the linearity principles of Chapter 2, Final Report, Volume 1 [1] .

It is possible to conceive of other methods of coding; for example, pulse width modulation with the signal information carried in the width of the pulse or, perhaps, pulse frequency modulation with the signal information carried in the frequency of the pulse transmission (Black [6], p. 30-36). These alternatives do not appear to be used, except in specialized applications (digital to analog conversion units, for example, Nelson [7]), and consequently will not be discussed here.

Sampling is not thought to be used in connection with a control system because of any advantage inherent in the sampling operation itself, but rather because of external reasons, e.g. the time sharing of equipment and the use of digital computers for control and compensation. An exception to this philosophy is the use of sampling devices with instrument servomechanisms which permits employment of highly sensitive error detectors (Marshall [8], p. 153).

It is not anticipated that over-all systems for use by the Air Force will receive digital input signals nor produce digital output signals, but that signals of this special nature will be present only within the control loop. As an example, a ground-to-air missile control system may well transmit data in digital form, but the desired missile angle of attack, to cite a variable to be controlled, will be initially in analog form and so will the actual angle of attack. Sub-systems of the main system may receive and produce digital signals, but these will be specialized components and must be dealt with as such. Thus systems with

either digital (i.e. coded) inputs, digital outputs or both digital inputs and outputs will not be considered here.

A sampled data system with continuous or analog input and output, when viewed from the input and output terminals, does not present a special problem due to the presence of the sampler as far as measurement of performance is concerned. In fact, the observer need never know that the system contains sampled or digital signals, as any peculiarity due to the presence of the sampler etc. will be observed at the output. Consequently a system can be considered satisfactory provided it can meet the input-output specifications placed upon it. The systems under consideration in this chapter must comply with the linearity principles, and it can be concluded, therefore, that all specifications recommended for use with linear continuous systems will be meaningful and shall be applied to sampled data systems.

The principal mathematical tools available for the analysis of sampled data system models are:

1. The z-transform (Ragazzini [9]) (Jury [10]), which can be made to yield a continuous function as the output, but which is valid only at the sampling instants.

2. The modified z-transform (Baker [11]) (Jury [12]), which yields the output at all instants of time at the cost of some algebraic complexity.

3. The so-called "state transition method" of analysis, which is possibly the most basic and has received attention in the literature as such (Gilbert [13]) (Kurzweil [14]) (Kalman [15]). It is more general in application than the z-transform but has not yet found general usage.

Use of the z-transform has the obvious disadvantage that the system
response between the sampling instants is left in doubt. An analysis on
this basis would fail to reveal oscillations that are entirely between
the sampling instants (Jury [16]), thus use of the modified z-transform
is necessary in systems where such responses are possible. The conditions
under which oscillations may occur and the methods of analysis in the
z-plane are, however, well known [12], [16], (Johnson [17]), (Schmidt[18])
and need not be detailed here. The important conclusion is that an output,
to which specifications can be applied, is available from a system mathe-
matical model.

The philosophy of Chapter 6 (Presentation of Data and System Perform-
ance Information), Final Report, Volume 1 [1] is also applicable here.
A system may meet all specifications, but it is desirable that a prospective
customer (e.g. the Air Force) be furnished with more details than are
presented by the system specifications alone.

This chapter is concerned with sampled data systems that have both
analog inputs and analog outputs and can be termed linear within the
principles of Chapter 2, Final Report, Volume 1. The restriction to
linear systems is consistent with the state of the art, i.e. any attempt
to apply specifications to nonlinear sampled data systems would require
considerable further research, which, while very important, is outside
the scope of this work. The restriction of the input and output quantities
to analog form thus excludes sub-systems that receive or transmit digitally
coded data. These sub-systems, e.g. a digital computer, are considered
to be specialized components and are not discussed here.

The specifications to be used are those recommended for use with
linear continuous systems, supplemented, as is dictated by the require-

ments of a particular system, by compatibility consideration peculiar to
sampled data systems as detailed in the next sections.

2.2 Recommended Specifications

Sampled data systems that are linear and time-invariant within the
principles of Chapter 2, Final Report, Volume 1 must be subdivided into
two classes. The sub-division is based upon the sampling device frequency,
ω_s, and the system bandwidth, BW.

Time varying, continuous systems are discussed in Chapter 3 of this
report, and the philosophy and principles discussed there can be extended
to time varying, sampled data systems with high sampling rates. The
problems that arise with time varying, continuous systems are multiplied,
however, when time varying sampled data systems with slow sampling rates
are considered. Extension of the chapter on time varying continuous
systems to this latter case is not recommended.

Sub-division one:

Systems with high sampling rate i.e. for which $\frac{\omega_s}{BW} \geqslant 10$

Sub-division two:

Systems with low sampling rate i.e. for which $2 \leqslant \frac{\omega_s}{BW} \leqslant 10$

The specifications recommended for linear continuous systems and
reproduced in Chapter 1 for reference are recommended for use in the
specification of all sampled data systems (i.e. both sub-divisions above)
whose output and input are available in analog form. In addition certain
"compatability considerations" must be considered.

Compatability considerations must be considered in all automatic
control systems where sub-systems of a larger system are constructed.
For example, impedance levels at the input and output of the system

must be compatible with the systems to which it is coupled and, in the case of a.c. control systems, the carrier frequency must be compatible. Such considerations are particularly important in sampled data systems.

The system sampling frequency may well be determined by factors external to the system or sub-system and may need to be specified as a compatibility specification for systems in both of the above sub-divisions.

Systems that fall in the second of the above sub-divisions can be expected to give inferior performance to those in the first sub-division, and consequently more care must be exercised with the specification of these systems. Compatibility considerations in addition to the sampling frequency already mentioned are: a) the amplitude, and the tolerance on this amplitude, of the harmonic content of the output to a sinusoidal input of fundamental frequency, and b) the maximum tolerable amplitude of the sampling ripple.

2.2.1 Discussion of Recommended Specifications

Analog output information is available for sampled data systems that are in the design stage and represented by mathematical models, and from systems that exist physically. The specifications already recommended for use with linear, continuous systems can be used, therefore, to assess the performance of sampled data systems. If the system response fits within the region of the output magnitude-time or magnitude-frequency space, as defined from the recommended specifications, the system is satisfactory.

Systems that include the sampling device within the loop may, however, exhibit behavior that can be attributed directly to the presence of this device. The design techniques available for this class of system, for example minimum finite settling time and zero

steady state error design ([4], p. 151), suggest response characteristics which should be controlled. The procedure whereby these additional quantities are controlled will be called "compatibility considerations". The presence of the sampling device is thus ignored for overall performance specification, but the peculiarities of the device are examined to ensure sub-system compatibility.* That is, the same specifications must be met whether or not the system contains a sampler.

One of the principal characteristics associated with a sampling device is the rate at which samples are obtained from the continuous data. The rate of sampling may well be a specification in itself dictated by circumstances outside the control of the designer. It is clear that the rate of sampling chosen will affect the performance of the system; in fact, as the sampling rate is increased, system performance will approach that of a continuous system [18] (Brown [19]). It is recommended, therefore, that sampled data systems be sub-divided according to the rate of the sampling device. First, those with a high sampling rate comprise sub-division one, where high sampling rate systems, as discussed Appendix A, have been defined as those systems where the sampling frequency ωs (ωs = $\frac{2\pi}{T}$ and T is the sampling period) is equal to, or more than, ten times the bandwidth of the system.

Mathematically:

$$\frac{\omega_s}{BW} \geqslant 10$$

ω s - sampling frequency

BW - system bandwidth

The greatest input frequency a system will be expected to experience

---

*It is not the intention of the authors to enter into a discussion at this point whether continuous systems are contained in the class of sampled data system or vice versa but rather to set satisfactory standards for the specification of either type system.

at its input terminals should be related by the designer to the system
bandwidth. The system bandwidth and the greatest input frequency expected
will therefore be considered synonymous in this chapter.

Systems that do not fall into the category covered by the above
restriction comprise the sub-division two. These latter systems are
defined in terms of the bandwidth and the sampling frequency by the
inequality:

$$2 \leqslant \frac{\omega_s}{BW} \leqslant 10$$

The lower limit is determined from the Nyquist Sampling Theorem, the
principle of which was first discussed in 1928 (Nyquist [20]).
Shannon [21] proves this theorem in a concise fashion and states the
principle as "If a function f(t) contains no frequencies higher than
W cps, it is completely determined by giving its ordinates at a series
of points spaced $\frac{1}{2W}$ seconds apart". The upper limit is determined from
Appendix A.

Systems that fall into the class where the sampling frequency can
be considered high will be specified by means of the specifications
recommended for linear, continuous systems and these specifications
will be sufficient input-output specifications.

The linear, continuous frequency domain specifications remain fully
meaningful for those sampled data systems that fall into the second sub-
division. A possible exception that should be mentioned is the assumption
that the system under consideration will possess low-pass filter charac-
teristics, and consequently the fundamental is the predominant component
in the output signal. If this assumption is invalid, the frequency
specifications, which are based on the concept that the output derived
from a sinusoidal input is of the same frequency as the input, begin to

lose their meaning. In such cases the same specifications can be applied to the input and output of fundamental frequency. In addition the amplitudes of the harmonic content of the output, together with the tolerance on these amplitudes, should be investigated as compatibility considerations.

The linear, continuous Time Domain specifications also remain fully meaningful when applied to the analog output data from systems in this second sub-division. Two characteristic difficulties associated with this class of system, namely inter-sampling ripple, introduced by the sampling device ([1] , p. 336), and hidden oscillations between the sampling instants [19] must be controlled by specifications. Care must be exercised when the output signal is derived from a z-transform analysis. Such an analysis, and the smooth curve through the sampling instants that it yields, may be misleading. Hidden oscillations and/or excessive inter-sampling ripples may go undetected.

The conditions under which oscillations, contained wholly between the sampling instants, will occur are well defined [16] ([2], p. 356) and can be avoided.

Inter-sampling ripple is introduced into the system by the higher frequency components generated during the sampling process. These components are attenuated, frequently, by system elements that exhibit low-pass filter characteristics, but some may remain at the output terminals. This inter-sampling ripple may not be troublesome during system transients but could be the only output after the transient has subsided ([2] , p. 338). Control of such a ripple is often essential.

Some control of this ripple will have been allowed for already by the Final Value of Error Specification, which limits the actual output.

to a region about the desired value of the output. The frequency domain specifications may also tend to limit the amplitude of the ripple. The frequency specifications are, however, based on input-output quantities whereas the ripple is produced internally. Direct control of the ripple amplitude may often be desirable.

It is recommended, therefore, that the maximum tolerable amplitude of the intersampling ripple be investigated as a compatibility consideration.

Performance Indicies were not recommended as performance specifications for linear, continuous systems and are not recommended for use with sampled data systems.

2.3 Data Presentation

It is recommended that Nyquist Diagrams and Root Locus Diagrams be used to display system data when describing system performance for the Air Force. The recommended methods are those recommended for use with linear, continuous systems. The diagrams will be supplied to the Air Force in addition to the performance specifications already discussed.

Performance specifications and their tolerance can be summarized in terms of time and frequency domain graphs as is indicated in Chapter 1. Reference has been made to the necessity for linearity checks and the principles of Reference 1 Chapter 2 indicate that a need may arise when these graphs should be presented for a number of input magnitudes. It is recommended, therefore, that actual time and frequency graphs taken for the system under consideration be presented for different input magnitudes as discussed in Reference 1, Sections 6.10 and 6.11.

Lastly the transfer function of the linear, continuous portion of the system is recommended for inclusion as system data since it expresses system characteristics in a concise way.

2.3.1 Discussion of Data Presentation Methods

Performance Specifications contain the information needed to evaluate a system in operation, as they describe the system on an input-output basis. It is often necessary, however, to consider additional factors less tangible then the numerical values of the specifications already recommended. The objective may be the evaluation of proposals and the selection of superior designs with regard to such factors as, for example, simplicity of design or sensitivity of parameter variation.

It is essential that this information (i.e. system data) be presented

in a form that will be familiar to those who have to evaluate the system. Furthermore, the form of data presentation should be of one or more specific forms so that the Air Force will have a common denominator for system comparison.

This philosophy follows that expressed in connection with linear, continuous systems. The principal analysis, (design), techniques available for use and currently used with sampled data systems will be enumerated with the purpose of selecting the most appropriate method or methods of data presentation.

2.3.2 The Routh-Hurwitz Criteria

These criteria cannot be applied to the characteristic equation of a sampled data system, which is in terms of the complex variable "s", associated with the Laplace Transform, as the equation is transcendental ([3], p. 522) ([4], p. 98). When the system characteristic equation is expressed as a function of the complex variable "z" associated with the z-transform where $z = e^{Ts}$, the criteria are not applicable either, as the transformation maps a horizontal strip of the left-hand half-plane, of the s-plane into the interior of a unit circle centered at the origin in the z-plane. Stability is now assured when the zeros of the characteristic equation are inside this circle and clearly the Routh-Hurwitz Criteria are not applicable.

The criteria can be applied, however, in a manner identical to that used with linear, continuous systems if the characteristic equation is expressed in terms of a complex variable w by means of the Mobius or Fractional Linear Transformation (Hille [22], p. 46):

$$z = \frac{1 + w}{1 - w}.$$

This transformation maps the interior of the unit circle, centered at the z-plane origin, into the left-hand half-plane of the w-plane. Stability is now assured if all the zeros of the characteristic equation in terms of w are in the left-hand half-plane of the w-plane and the criteria can now be applied. Gain margin information is available as a result of this analysis in the w-plane and the results can be transferred back through the transformations to the s-plane.

The labor involved in this operation may well become extensive and the information that results is only Gain Margin, which is inadequate for system evaluation. The Routh-Hurwitz Criteria is not recommended for data presentation.

2.3.3 The Schur-Cohn Criterion

The criteria just described above are able to detect the presence of roots with positive real parts of a polynomial expressed in terms of a complex variable. In the case of sample data systems the characteristic equation must be examined for roots that lie outside the unit circle in the z-plane. The Schur-Cohn criterion ([2], p. 238) comprises an elegant test for the determination of such roots. The information obtained from this test is, however, restricted to Gain Margin, as was the case with the Routh-Hurwitz Criteria, and this information is not sufficient for system evaluation. The Schur-Cohn Criterion is not recommended for use as a method of data presentation.

2.3.4 The Bode Diagram

The open loop transfer function of a sampled data system, when written in terms of the Laplace Transform complex variable "s", cannot be expressed as the ratio of finite polynomials. The Bode Diagram, as a logarithmic plot of magnitude against frequency, thus loses the im-

portant advantages it has for linear, continuous systems; that is, ease

of construction and identification of time constants. The familiar

linear continuous equalizer procedures, should one wish to use such an

equalizer, are not valid because the continuous transfer function of an

equalizer cannot be added directly to the open loop plot on the diagram,

as is the case with linear, continuous systems. The use of pulsed data

equalization can be effected on the Bode Diagram, but the technique is

difficult to apply ([2], p 432).

It is possible to make use of the familiar linear, continuous

design techniques on the Bode Diagram by approximating the open loop

transfer function ([4], p. 124), but the approximation is inaccurate

for low sampling frequencies where accuracy is most desirable.

An alternative approach is to transform the open loop transfer

function to the z-plane and then to the w-plane. The Bode Diagram

technique is now directly applicable, but physical reality has been

lost by the sequence of transformations. The principal advantage of

this method of data presentation, that of insight into the system

capabilities, is thus lost also.

The Bode Diagram is not recommended as a method of Data Presentation

for sampled data systems.

2.3.5 The Nyquist Diagram

The diagram, when used in connection with sampled data systems,

is constructed and can be used in a manner similar to that for linear,

continuous systems ([3], section 9.6).

The diagram is a polar plot of the magnitude and the phase of the

open loop transfer function as a function of frequency. Where $G(j\omega)$

is the continuous forward transfer function, the transfer function in-

cluding the sampling device becomes:

$$G^* (j\omega) = \frac{1}{T} \sum_{-\infty}^{+\infty} G\ j(\omega + n\omega_s) \qquad \frac{2\pi}{T} = \omega_s = \text{sampling frequency}$$

or in terms of the variable z merely G(z).

The transfer function can be plotted as the Nyquist diagram, approximately from the first expression using only the first few terms of the series. It may also be plotted exactly, directly from G(z), recalling that points on the $j\omega$- axis of the s-plane correspond to points on the unit circle centered at the origin in the z-plane.

M-circles (Chestnut [23], section 9.2) can be constructed and are fully meaningful ([2], p. 413) with the usual restriction that the system must be one of unity feedback.

In conclusion, the diagram provides the same information as it did in the case of linear, continuous systems. It was recommended there for use in the presentation of data and, therefore, is recommended for data presentation with sampled data systems.

2.3.6 Root Locus Plot

The root locus diagram can also be extended to display the movements of closed loop poles, as a function of a system parameter (Jury [24]) (Mori [25]), with sampled data systems. In the s-plane there are infinitely many open-loop singularities, but loci can be drawn. The root locus can be drawn more simply in the z-plane, with the disadvantage that position with respect to the unit circle and the origin of the z-plane is of importance, rather than the more familiar concept of position with respect to rectangular axes in the s-plane.

The transient response is characterized by the position of the closed loop poles in the z-plane, and hence by the position of the root

loci. The lack of familiarity with the geometry associated with the z-plane can be overcome to some extent by, for example, constructing contours of peak overshoot [12] related to a particular system on the z-plane, together with the root locus plot.

Root loci can also be drawn in the w-plane, using the transformation already described, but, as mentioned before, physical reality is lost.

Root locus diagrams are recommended for use by the Air Force for data presentation. The diagrams should be plotted as a function of one or more of the parameters that are of interest and in the z-plane.

The diagrams recommended for use by the Air Force are the Nyquist Diagram and the Root Locus Diagrams in the z-plane. These diagrams are those recommended for Air Force use in connection with linear, continuous systems. A factor used to aid in the selection of these diagrams in Final Report, Volume 1 [1] was the possible extension to nonlinear systems. It is fortunate that the advantages in using these diagrams with sampled data systems lead to their recommendation here, and thus the possibility of using these diagrams with all systems exists.

The Nyquist Diagram as drawn for a sampled data system in the s-plane can be compared directly with a similiar diagram for a continuous system. Consequently continuous and sampled systems can be compared by means of a Nyquist Diagram. This comparison cannot be made as easily when the Root Locus Diagram is used. The continuous system diagram will be drawn in the s-plane and the sampled systems diagram in the z-plane and the appearance of the two diagrams is quite different. The technique in the z-plane for sampled systems is the same, however, as the technique in the s-plane for continuous systems. The z-plane technique is common in

the literature and the transient response information is available from the digram (Jury [26]). The z-plane root locus is thus recommended in order to provide this transient information despite the difficulties with such diagrams.

CHAPTER 3

THE SPECIFICATION OF LINEAR TIME VARIABLE
PARAMETER SYSTEMS

Abstract

Linear time variable parameter (LTVP) systems are defined and sub-
divided into those systems with fast or slow variations and/or large or
small variations. The methods of analysis of such systems are reviewed,
and the following recommendations are made.

Specifications

1) Time Domain Specifications

    (a) LTVP systems with fast variation of parameters.

        Simulated unfrozen system step function responses should
all lie within a prescribed envelope. Whenever possible, the
actual system response should be obtained.

    (b) LTVP systems with slow variation of parameters.

        Simulated or actual frozen or unfrozen system step function
responses should all lie within a prescribed envelope.

2) Frequency Domain Specifications

    (a) LTVP system with fast variation of parameters.

        Frequency domain specificationsare not recommended.

    (b) LTVP system with slow variation of parameters.

        The family of frequency response curves of the system
frozen at different instants should all lie within a prede-
termined envelope.

Data Presentation

It is recommended that the region of variation of closed

loop poles of the frozen system be exhibited on the complex

plane. Thus, for example, if the only varying parameter is an

open loop gain then the region of variation of the closed loop

poles will correspond to the root loci over the total range of

variation of gain.

It is also recommended that a family of Nyquist diagrams

corresponding to the system frozen at different instants be

displayed in the case of systems with slow variations of parameters.

3.1 Introduction

Time variable parameter systems occur more often in practice than

one with a fair knowledge of control system theory might suspect. In

fact, the statement that most practical systems are nonlinear and time

variable is not very far from the truth. However nothing of value is

known at the present moment regarding the analysis and synthesis of

general nonlinear time variable parameter systems. Considerable research

effort is being expended to solve certain facets of this problem,

particularly by workers involved with self-adaptive systems which are,

in general, nonlinear and time varying. Inasmuch as it is highly

desirable to be able to specify performance criteria for general

systems of this kind, the state of the art at the moment leaves so much

to be desired that it does not appear to be feasible now.

This particular report is restricted to the special case of linear

time variable parameter systems (henceforth referred to as LTVP systems).

LTVP systems are also sometimes referred as nonstationary linear systems.

There is not a very great loss of generality by this restriction since

one will soon see that not very much more is known regarding possible
means of specifying LTVP systems than nonlinear time variable systems.
It is possible to approximately describe a number of practical control
systems in such a fashion that they may be considered to be LTVP systems.
For example, a missile subjected to thrust due to fuel burnout may be
considered as a LTVP system by assuming that the burnout rate is a
constant. This latter assumption is, of course, quite reasonable in
general.

The analysis of LTVP systems is also important sometimes from the
viewpoint of study of special characteristics of certain nonlinear
systems. For example, the study of the periodic solutions of certain
forced nonlinear systems (for example, systems being described by
equations of the Duffing type) resolves into studying equations similar
to the ones which govern LTVP, such as Mathieu's equation (Stoker [27]).

The study of the analysis and synthesis of LTVP systems is a rather
interesting, important and fascinating research topic at the present
time. However the techniques available at the moment leave much to be
desired. Because of this, the complete specification of performance
criteria for LTVP systems presents a rather formidable problem that is
unsolved as yet.

Before discussing the specific problem of how to describe the per-
formance of LTVP systems, let us first consider the general LTVP system
and the several known methods of analysis of them.

Definition 3.1

A general linear lumped parameter time variable system is described by a differential equation of the following type.

$$a_n(t) \frac{d^n y}{dt^n} + a_{n-1}(t) \frac{d^{n-1} y}{dt^{n-1}} + \ldots\ldots\ldots + a_1(t) \frac{dy}{dt} + a_o(t) \, y =$$

$$b_m(t) \frac{d^m x}{dt^m} + b_{m-1}(t) \frac{d^{m-1} x}{dt^{m-1}} + \ldots\ldots\ldots + b_1(t) \frac{dx}{dt} + b_o(t) \, x \qquad (3.1)$$

In equation (3.1), y is the output and x the input of the system. The coefficients a_n, $a_{n-1}\ldots\ldots b_m$, b_{m-1},$\ldots\ldots$are functions of time alone and are assumed to be piecewise continuous over any finite subinterval. Furthermore, it is assumed that $a_n(t)$ does not vanish at any point in the interval of interest $a < t < b$.

Equation (3.1) may be represented in the following operational form

$$L(D,t) \, y = K(D,t) \, x \qquad (3.2)$$

where

$$L(D,t) = a_n(t) \, D^n + \ldots\ldots + a_1(t) \, D + a_o(t) \qquad (3.3)$$

$$K(D,t) = b_m(t) \, D^m + \ldots\ldots + b_1(t) \, D + b_o(t) \qquad (3.4)$$

and

$$D = \frac{d}{dt}$$

The quantity n is defined as the order of the system and in general, for a physical system $n > m$.

LTVP systems obey the powerful superposition theorem, due to the fact that equation (3.1) is a linear differential equation. At first glance, this fact may lead one to believe that the analysis of LTVP systems is not very different from the well known methods of analysis of linear time invariant systems. Unfortunately, this is far from the truth. Very few

methods of analysis exist for LTVP systems, and even the methods that are developed are applicable only to specific types of systems.

A linear system is completely specified by its impulse response $h(t,t_1)$. Note that in general the impulse response is a function of two instants of time, t the time of observation of the output and t_1 the time of application of the input. In the case of time invariant systems, the impulse response is a function only of the "age variable" $t-t_1 = \tau$. It is this latter property which makes the analysis of linear time invariant systems mathematically tractable. Moreover, the dependence of the impulse response on one variable is the reason that any significant meaning may be attached to time-domain and transient response specifications for linear time invariant systems. The transient response characteristics, such as the impulse response or the step function response, for a linear time invariant system can be obtained from a complete knowledge of the response of the system to any one transient input applied at any instant of time. However, the transient response of a LTVP system implies, theoretically, knowledge of an infinity of responses obtained by application of the transient input at different instants of time. This makes the specification of LTVP systems in terms of transient response such as, for example, step function response, generally meaningless. Frequency domain specifications have less meaning for LTVP systems since a harmonic input to a LTVP system does not in general result in a harmonic or even periodic output.

It is evident from the literature in the area of differential equations (Ince [28]) (Bellman [29]) that considerable effort has been spent in determining the stability characteristics of linear, time variable differ-

ential equations. This is evidenced, for example, by the elegant Floquet's theory ([27], p. 193) in connection with Hill's equation and Mathieu's equation. This theory is applicable to a particular type of second order differential equation and despite its elegance, is useless as far as performance specifications of LTVP systems are concerned. It should be noted that it is possible to discuss the stability of a LTVP system only in the absence of any inputs, whereas the performance specifications of a system in general are based on some form of input to the system.

In this connection it may be worthwhile to precisely define the concept of stability for a LTVP system. The following definitions are equivalent.

Definition 3.2

A LTVP system is defined as stable if the complementary solution (transient solution) associated with its differential equation, of the form of equation (3.1) identically approaches zero when time increases beyond all bounds for any arbitrary initial conditions.

Definition 3.3

A LTVP system is defined as stable if its impulse response $h(t, t_1)$ is absolutely integrable over the infinite range of t for all values of t_1. (Zadeh [30], p. 403)

Note that the impulse response $h(t,t_1)$ may be obtained from equation (3.2) as the solution of

$$L(D,t) \, x h(t,t_1) = K(D,t) \, x \, \delta(t-t_1) \tag{3.5}$$

where

$\delta(t - t_1)$ is the Dirac-delta function.

For LTVP systems, both the definitions imply quasi asymptotic
stability of the system and not necessarily quasi uniform stability
(Antosiewicz [31], p. 147). However, for linear time invariant systems,
a stable system on the basis of definitions 3.1 and 3.2 implies a
uniformly asymptotic stable system.

It is evident that these definitions of stability for LTVP systems
are not particularly useful for control system applications. For example,
the transient response of a "stable" LTVP system (in the sense of de-
finitions 3.1 and 3.2) might exceed a safe value (possibly resulting in
a destruction of the system) at some instant of time, despite the fact
that the response appeared well behaved for a reasonable length of time
after application of the input. This is a problem that is not encountered
in a linear time invariant system. For example, the first maximum of the
step function response (corresponding to the overshoot) is the absolute
maximum for a stable linear time invariant system.

It is sometimes useful to define stability on the basis of uniformly
asymptotic stability.

Definition 3.4

     Any uniformly bounded input should give rise to a uniformly

     bounded output in a uniformly asymptotically stable LTVP system

     (Kalman [32], p. 379)

According to Massera's theorem (Massera [33], p. 204) in order for
the LTVP system to be totally stable, i.e. stable for every bounded in-
put, it should be uniformly asymptotically stable.

For certain control system applications, the stability of a LTVP
system may be specified on a short time basis. For example, a LTVP
system may be defined to be short time stable if the response to a

specific type of input such as a step remains within certain predetermined bounds for a specified interval of time after the application of the input. The behavior of the system outside this interval is of no consequence.

It is realized that stability is a rather important and interesting characteristic in the analysis of systems. However, the response of the systems to certain specific inputs is more important in specifying the performance of the systems. Various authors have proposed specific methods for determining the response of particular LTVP systems, (Gerardi [34]), (Kirby [35]), (Kirby [36]), (Brodin [37]), (Bennett [38]), (Karamyshkin [39]), (Desoer [40]). This suggests the possibility of trying to classify LTVP systems on the basis of being amenable to the various special methods. This raises a rather difficult problem which shall be discussed later.

3.2 Methods of Analysis

To determine the feasibility of classifying LTVP systems, the various methods that are available at the moment for studying their stability or obtaining their responses are listed below with proper references. Brief explanations of the methods are given where necessary. It should be noted that there is a certain amount of overlap between the methods.

3.2.1 Analysis Using Classical Differential Equation Theory [28]

From the theory of differential equations it is known that a unique solution of equation (3.1) exists and consists of two parts. They are:

(a) The complementary solution, which is the solution of the homogeneous equation

$$L (D, t) u = 0 \tag{3.6}$$

Thus, if u_1, u_2, ... u_n are the n distinct* solutions of equations (3.6), then the solution

$$u(t) = c_1 u_1 + c_2 u_2 + \ldots + c_n u_n \qquad (3.7)$$

containing n arbitrary constants is the complementary solution.

(b) The particular solution, which is any solution $y_0(t)$ which satisfies the nonhomogeneous equation (3.1). The complete solution of (3.1) is then given by

$$y(t) = y_0(t) + u(t) \qquad (3.8)$$

It is generally difficult to obtain the complete solution of a differential equation of the form (3.1) for any arbitrary input. Note from the definition of stability that the complete stability information for the system is contained in the complementary solution, equation (3.7)

3.2.2 Analysis of LTVP Systems by Matrix Methods [29], (Pipes [41])

The matrix analysis of LTVP systems is not significantly different from the classical method of analysis. The only advantage is that the notation is simple enough to prevent one from getting involved in the algebra associated with the classical method.

---

* A sufficient condition for the linear independence of $u_1, u_2, \ldots u_n$ (which is the same as saying that they are distinct) is that the Wronskian of the functions u_1, u_2, ... u_n should not be identically zero. This means that

$$\begin{vmatrix} u_1 & u_2 & \ldots & u_n \\ u_1' & u_2' & \ldots & u_n' \\ u_1^{(n-1)} & & \ldots & u_n^{(n-1)} \end{vmatrix} \neq 0$$

Here

$$u^{(k)} \triangleq \frac{d^k u}{dt^k}$$

For example, if one were just interested in the stability problem*, the nth order homogeneous differential equation, equation (3.6), may be reduced by a suitable transformation to a set of n first order differential equations of the form

$$\frac{d\,Y}{d\,t} = A(t)\,Y \tag{3.9}$$

where Y is a column vector with n elements and A(t) is an n x n square matrix whose elements are functions of time in general.

The homogeneous equation associated with the system, equation (3.6), may be reduced to the form of equation (3.9) by defining new variables for the output and its derivatives, as follows:

Let

$$y = y_1 \tag{3.10}$$

and

$$
\left.
\begin{aligned}
\dot{y}_1 &= y_2 \\
\dot{y}_2 &= y_3 \\
\dot{y}_{n-1} &= y_n
\end{aligned}
\right\} \tag{3.11}
$$

then from (3.6)

$$\dot{y}_n = \frac{-a_o(t)}{a_n(t)}\,y_1 - \frac{a_1(t)}{a_n(t)}\,y_2 - \cdots \frac{-a_{n-1}(t)}{a_n(t)}\,y_n$$

$$= -c_o(t)\,y_1 - c_1(t)\,y_2 - \cdots - c_{n-1}(t)\,y_n \tag{3.12}$$

where

$$c_i(t) = \frac{a_i(t)}{a_n(t)} \tag{3.13}$$

Equations (3.11) and (3.12) may be combined and written in the form of equation (3.9) by defining Y (t) and A (t) suitably as follows

---

* This information is completely contained in equation (3.6).

Let

$$Y(t) = \begin{Bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{Bmatrix} \qquad\qquad (3.14)$$

and

$$A(t) = \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ & & & & & \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ -c_0 & -c_1 & -c_2 & \ldots & -c_{n-2} & -c_{n-1} \end{bmatrix} \qquad (3.15)$$

The y_i's represent the so called <u>state variables</u> of the system. In general, any independent linear combination of state variables is also a state variable. The n-dimensional space with each state variable represented along a co-ordinate axis is called the state space.

Theoretically, the stability of the time varying system is completely determined by the matrix $A(t)$. However, at the present moment, there appears to be no generalized theory which can be applied to every matrix $A(t)$. The stability of certain systems yielding specific types of matrices $A(t)$ has been investigated in reference [3].

The matrix method could also be used for obtaining the response of some specific LTVP systems. However, this method does not possess any distinct advantage over the classical method.

<u>3.2.3 Frequency Analysis Approach [30], (Zadeh [42], [43], [44], [45])</u>

This is essentially an attempt to extend the familiar concept of

poles and zeros to LTVP systems. The theory developed in this case closely parallels the system transfer function concept for linear time-invariant systems. However, the application of this method to any but the most trivial cases is rather difficult. However, this method is not without advantages. With the transfer function concept for LTVP systems, one can intuitively picture poles and zeros of the LTVP system wandering in the complex plane as functions of time. The question of stability can immediately be settled for some intuitively obvious cases [38].

The essence of the method hinges on obtaining the so called system function $H(j\omega;t)$ which is an integral transformation quite analogous to Fourier transformation of the weighting function $h(t,t_1)$, defined by equation (3.5). The system function is defined by

$$H(j\omega;t) = e^{-j\omega t} \int_{-\infty}^{\infty} w(t,\xi) \, e^{j\xi} \, d\xi \qquad (3.16)$$

In general it is very difficult to obtain the system function. Zadeh ([42], p. 295) shows that the system function satisfies the non-homogeneous linear partial differential equation with complex coefficients which are functions of time of the form

$$\frac{1}{n!L} \frac{\partial^{n}L}{\partial(j\omega)^{n}} \frac{\partial^{n}H}{\partial t^{n}} + \dots + \frac{1}{L} \frac{\partial L}{\partial(j\omega)} \frac{\partial H}{\partial t} + H = \frac{K}{L} \qquad (3.17)$$

where L and K in equation (3.17) are obtained by replacing D by $j\omega$ in equation (3.3) and (3.4).

It is seen that Zadeh's frequency transformation is equivalent to going from the unsolvable equation (3.1) to a more difficult equation (3.17).

Zadeh also points out the intuitively obvious case of a LTVP system with slow variation of parameters, in which case the first approximation to the system function at any instant is the same as the transfer function

obtained by freezing the parameters at that instant. This is then defined as the frozen system function.

Zadeh also mentions the interesting possibility of the bifrequency transformation. This is essentially the Fourier transformation of the system function defined by equation (3.16) where the variable of transformation is t. No particular use of the bifrequency function of a system is known at this time.

In summarizing, it is felt that the frequency analysis approach transforms one unsolved problem in the time domain to another unsolved problem in the frequency domain.

3.2.4 The Transform Method([28], Chap. 8), (Aseltine [46])

In this particular method, an equation of the form (3.1) is solved by defining a suitable integral transformation, such that when both sides of equation (3.1) are operated on by this particular transform, a mathematically tractable equation results. This method gets quite complicated even for a second order system [45].

Aseltine considers the solutions of a second order LTVP system of the form

$$a(t) \, q'' + b(t) \, q' + d^2 q = e(t) \tag{3.18}$$

where the primes refer to differentiation with respect to time.

Aseltine seeks an integral transformation of the form

$$L\left[q(t)\right] \triangleq Q(\xi) = \int_0^t q(t) \, h(\xi, t) \, dt \tag{3.19}$$

where $h(\xi, t)$, the kernel, is a function of time and of the transform variable ξ, and $Q(\xi)$ is called the transform of $q(t)$.

It is now required that the application of the transformation (3.19)

to the LTVP system equation (3.18) results in

$$f(\xi)Q(\xi) + d^2Q(\xi) = E(\xi) + [\text{terms involving initial conditions}]$$ (3.20)

where $f(\xi)$ is an arbitrary function of the transform variable.

By redefining the kernel to include a function $g(t)$, which will make the differential operator of (3.12) self adjoint, one obtains

$$h(\xi,t) \triangleq g(t) \, xk \, (\xi,t)$$ (3.21)

it is shown that

$$g(t) = \exp \left[\int \frac{b(t) - a'(t)}{a(t)} \, dt \right]$$ (3.22)

and $k(\xi,t)$ satisfies the homogeneous partial differential equation

$$a(t)\frac{\partial^2}{\partial t^2} k(\xi,t) + b(t)\frac{\partial}{\partial t} k(\xi,t) - f(\xi) k(\xi,t) = 0$$ (3.23)

The kernel function of the integral transformation depends on the coefficients of the homogeneous differential equation (3.23). Thus, use of this method, even for a simple second order LTVP system, involves constructing tables of transforms and their inverses for each particular set of coefficients. It is seen that the work involved is monumental.

It can be shown that in the special case when the coefficients are constants (corresponding to a time invariant system), the suitable integral transform is the familiar Laplace transform.

3.2.5 Application of the Second Method of Liapunov [31], [32], [33]

(Malkin [47]) (Szego [48])

The second method of Liapunov is useful only for determining the stability of a LTVP system. No information can be obtained regarding the response of the system to any inputs.

The following exposition of Liapunov's second method essentially follows reference [48]. In order to present the theorem, the following

definitions are necessary.

Definition 3.5

A real scalar function $V(Y,t)$ is called positive semidefinite if

$$V(o,t) = o$$

and

$$V(Y,t) \geqslant o$$

(3.24)

Note that Y refers to the state variables vector of the system. The vector of state variables Y is defined by equation (3.14).

Definition 3.6

A real scalar function $V(Y,t)$ is called negative semi-definite if : $-V(Y,t)$ is positive semi-definite.

Definition 3.7

A real scalar function $V(Y,t)$ is called positive definite in the Liapunov's sense if

$$V(o,t) = o$$

and $V(Y,t) \geqslant w(Y)$

where $w(Y) > o$ for $y_i \neq o$

(3.25)

and $w(o) = o$

Definition 3.8

A real scalar function $V(Y,t)$ is called negative definite in Liapunov's sense if $- V(Y,t)$ is positive definite.

Liapunov's theorem

If for $t \geqslant t_o$ there exists a real scalar function $V(Y,t)$ in the neighborhood of the origin, $V(Y,t)$ being continuous and possessing continuous partial derivatives with respect to y_i and t, and satisfying some of the following conditions.

1. $V(Y,t)$ is positive definite in Liapunov's sense for $t > t_o$

2a. $\frac{dV}{dt}$ is not positive in some region S around the origin of the phase space for $t > t_o$.

or   2b. $\frac{dV}{dt}$ is negative definite in Liapunov's sense in S for $t > t_o$.

2c. $\frac{dV}{dt}$ is positive definite in Liapunov's sense in S for $t > t_o$.

3. $\displaystyle\lim_{\|Y\| \to o} V(Y,t) = o$ uniformly on t, for $t > t_o$.

where $\|Y\|$ refers to the Euclidean Norm of the vector Y.

4. $\displaystyle\lim_{\|Y\| \to \infty} V(Y,t) = \infty$ uniformly on t, for $t > t_o$.

Then the trivial solution $Y = 0$ corresponding to the origin of the state space is:

a. <u>Stable</u> in S if the conditions 1 and 2a are satisfied.

b. <u>Asymptotically stable</u> in S if 1 and 2b are satisfied and either $A(t)$ is bounded or there exists a real scalar function $\sigma(t)$ which is defined, continuous and increasing for $t > t_o$, with $\sigma(t_o) = o$ such that

$$\frac{dV}{dt} \leqslant -\sigma\left\{V(Y,t)\right\} \quad \text{for every } t > t_o.$$

c. <u>Uniformly Asymptotically stable</u> in S if 1, 2b and 3 are satisfied.

d. <u>Unstable</u> in S if 1, 2c and 3 are satisfied.

The crux of the Liapunov's second method lies in obtaining a suitable V function which will yield useful answers. Except for the trivial case of a linear time invariant system (for which the Routh criterion could be used to determine stability), there is no general method for determing a suitable V function.

3.2.6 Analysis by Simulation (Matyash [49]), (Laning [50])

This appears to be the most fruitful method for comparing LTVP systems. This essentially means that the system is simulated on the analog or digital computer. The response of the system to any desired input may then be determined. Here actual operation records of equivalent systems could be valuable.

3.2.7 Discussion of the Six Methods

The six methods for analyzing LTVP systems discussed so far have a common and serious disadvantage as far as performance specifications are concerned. The starting point for all these methods is the assumption that the time varying system can be mathematically described to a fair degree of accuracy. This assumption presupposes that it is possible to experimentally determine the mathematical description of a piece of hardware. This is a severe assumption from a strictly theoretical viewpoint. There is no known method whatsoever of obtaining the differential equation governing a piece of hardware by any means, experimental or otherwise, even if there is a priori knowledge regarding the linearity of the system, if the system happens to have time varying parameters. The last statement is true only if the system is treated from a pedagogical viewpoint of a black box with an input and output with no means of knowing what is inside the box. However this is rarely true in practice where it is often possible to estimate the equations governing the system with a fair degree of accuracy.

It is apparent from the discussion of the first five methods that analytical determination of the response of LTVP systems in any but the most trivial cases is a very difficult matter. However the simulation method could be used in almost all the cases where it is possible to

describe the system mathematically.

It is also evident from close examination of the first five methods outlined here for analyzing LTVP systems that this area warrants considerable research before any great progress can be expected in analyzing and synthesizing them. Despite the fact that the stability of LTVP systems is a rather interesting and intriguing problem and a number of researchers are working on this facet, it is felt that more efforts should be concentrated on determining approximate, if not exact, methods of obtaining responses of LTVP systems to specific inputs. The solution of the latter problem may be the answer to the problem of specifying LTVP systems.

3.3 Recommended Specifications

It is seen that the general problem of specifying LTVP systems is not an easy one to solve. However it is possible to classify certain LTVP systems so that some of the linear specification in reference [1] are valid for them.

It is difficult to obtain any specification for a LTVP system on the basis of absolute stability as defined earlier. However, for certain applications, it may be necessary to specify that the system should be asymptotically stable. For certain other applications (example: a control system for a "short life" missile) a short time stability specification may be sufficient.

For purposes of performance specifications, the following definitions are made for LTVP systems.

Definition 3.9 Fast and Slow Variation of Parameters

A LTVP system will be defined as fast varying if the <u>maximum rate of change</u> of any closed loop parameter (for example a closed loop pole

or zero) exceeds a predetermined value x per cent per second. A
typical value for x is unity. If the rate of variation is less than
x, the LTVP system will be determined as slowly varying.
Definition 3.10 Large and Small Variation of Parameters.

A LTVP system will be defind as having large variation of para-
meters if the maximum change in any closed loop parameter exceeds y
per cent. A typical value for y is 10. If the maximum change is
less than y, then the LTVP system will be defined as having small
variation of parameter.

The definitions here are made on the basis of variation of closed
loop parameters. The variation of open loop parameters is of no consequence.

An example of a system with large, slowly varying parameters is an
aircraft starting its flight with a full load of fuel. The initial
mass of the aircraft and fuel is comparatively large, and a mass change
of 20% due to the fuel being used up over a period of 5 to 6 hours is
typical. Here we see that the rate of change of mass is small whereas
the change is mass is large.

3.3.1 Time Domain Specifications

(a) LTVP systems with fast variations of parameters.

It is recommended that for acceptance a family of step function re-
sponses of the simulated "unfrozen system" satisfy the transient response
envelope specification [1]. The dimensions of the envelope shall depend
on the applications. Whenever possible, it is recommended that the step
function responses of the actual system be obtained. It is also recommended
that the mathematical description of the simulated system be furnished
along with the step responses.

(b) LTVP systems with slow variation of parameters.

It is recommended that for acceptance a family of step function responses of the simulated or actual, frozen or unfrozen system satisfy the transient response envelope specification, the dimensions of the envelope again depending on the applications. Whenever possible, it is recommended that the response of the actual system be used. If the system has been simulated to obtain the response, it is recommended that the mathematical description of the system be furnished.

3.3.2 Frequency Domain Specifications

Frequency response does not have any significant meaning in the case of a general LTVP system. This is because a harmonic input to a LTVP system may not even result in a harmonic output. This point is illustrated by the following simple example.

Consider the system shown in Fig. 3.1. Let e_i be the sinusoidal input $E \sin \omega t$ and e_o the output.

It is seen that

$$e_o = iR_1 \qquad \text{and}$$

$$e_i = \left[R_o + r f(t) + R_1 \right] i$$

Hence

$$e_o = \frac{R_1}{R_1 + R_o + rf(t)} \, e_i$$

$$= \frac{R_1}{R_1 + R_o} \frac{1}{1 + \frac{r}{R_1 + R_o} f(t)} \, ei$$

(3.26)

It is easy to see that equation (3.26) is similar to equation (3.1) and hence the system shown in Fig. 3.1 represents a LTVP system. Equation (3.26) may be rewritten as follows:
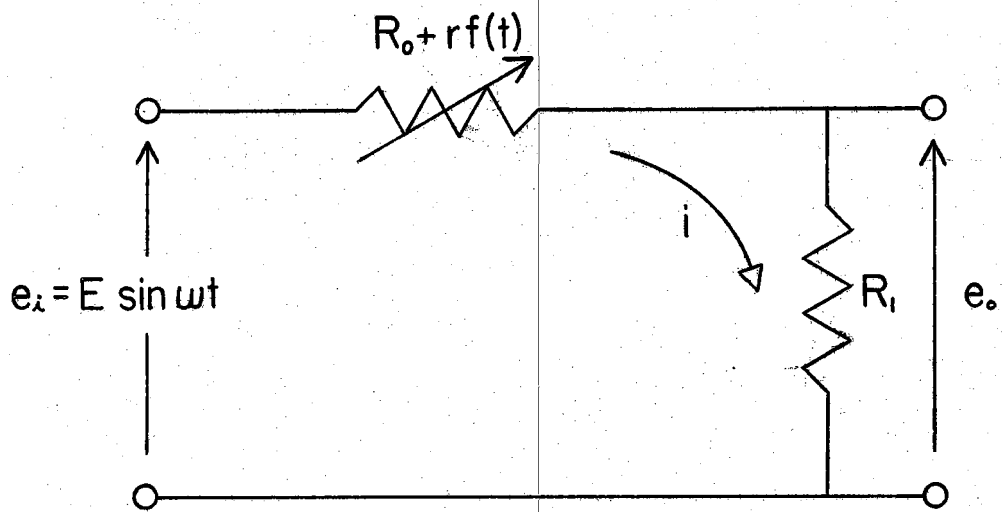
Figure 3-1

Simple Electrical Network with

a Time Varying Parameter

$$e_o = \frac{R_1}{R_1 + R_o} \left[1 - \frac{r}{R_1 + R_o} f(t) + \frac{r^2}{(R_1 + R_o)^2} f^2(t) - \dots \right] E \sin \omega t \qquad (3.27$$

It is seen that in any but the most trivial case corresponding to $f(t)$ = constant, the right hand side of equation (3.27) is not sinusoidal.

Frequency domain specifications are not recommended for systems with fast variation of parameters. For systems with slow variation of parameters, the "envelope specification" [1] on the frequency response of the "frozen" system is recommended. The dimensions of the envelope should depend on the applications.

The general state of the art of specifying LTVP systems is schematically shown in Fig. 3.2.
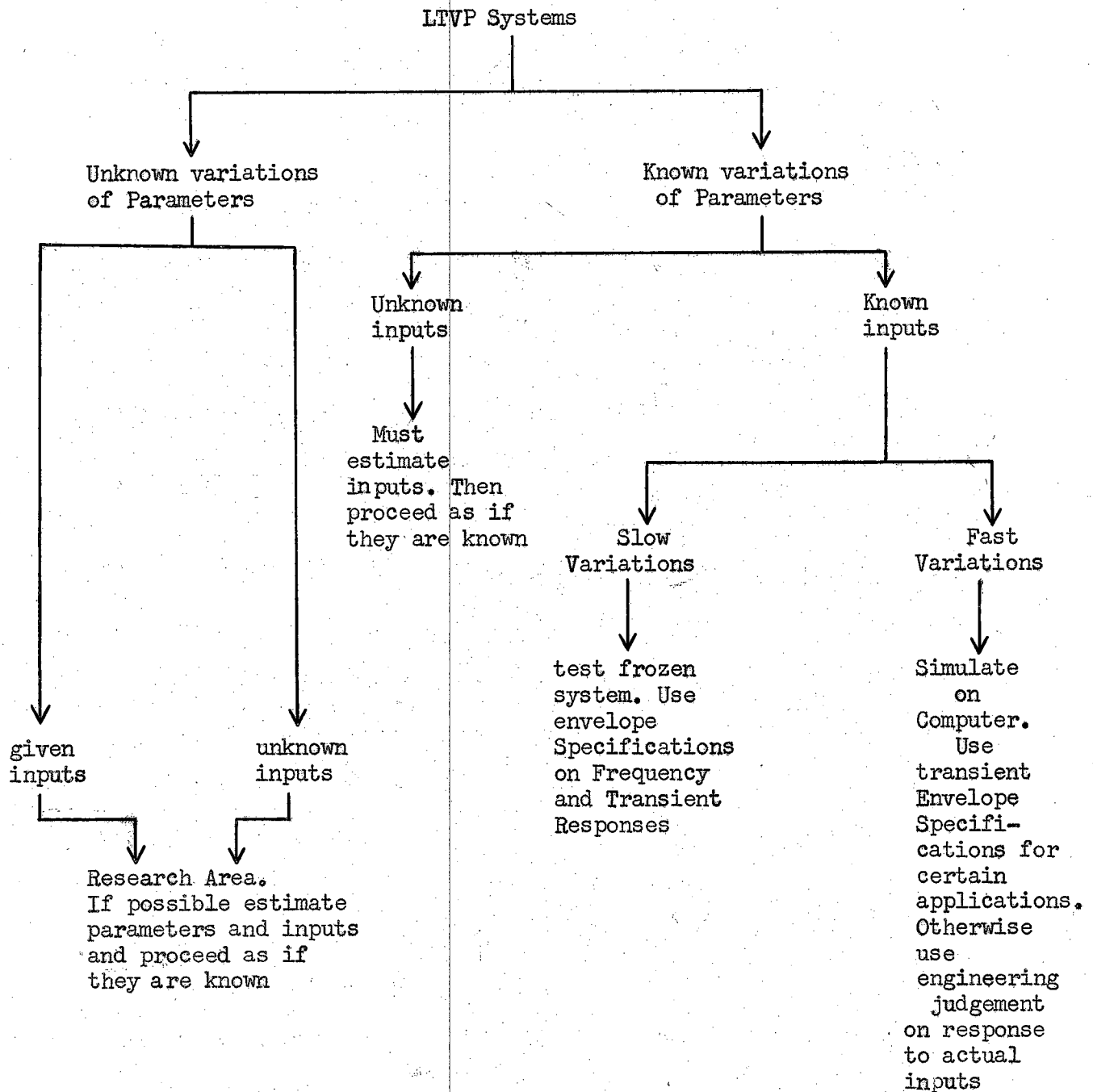
FIGURE 3.2 Schematic Representation of the LTVP
            Specification Problem.

CHAPTER 4

PERFORMANCE INDICES

Abstract

This study was undertaken to determine whether or not Performance

Indices should be used to evaluate and specify control systems. It is

recommended that they not be used at this time by the Air Force for the

stated purpose.

A performance index is defined and detailed discussions are pre-

sented for the various performance indices. Analytical methods for

evaluating performance indices are presented.

4.1 Introduction and Recommendations

A Performance Index* (Figure of Merit) has been defined by Anderson,

et al ([51], p. 182) as: "Some mathematical function of the measured re-

sponse, the function being chosen to give emphasis to the system specifi-

cations of interest." Ideally for evaluation, a Performance Index is a

single number in which a designer attempts to place his engineering

judgment on the overall excellence of a system. The Performance Index

may be chosen so that only one or a few system properties affect its

value. Or, it may be chosen so that it is a function of all the important

properties of a system's response. This second type of Performance Index

is the one of primary concern in this work. It is realized that there

are many other criteria, such as reliability, size, weight, cost, etc.

---

*The terms performance index and figure of merit are used interchangeably.
Most authors use the term Figure of Merit, but the term Performance Index
will be used here because a Figure of Merit is usually a quantity to be
maximized, whereas almost all the criteria here included are to be mini-
mized.

which must be taken into account in selecting a system. Bellman [52], for example, discusses a more general performance index. However, these are outside the scope of this project. The performance indices covered in this report consider only system response.

Control engineers have been interested in Performance Indices for over a decade. This interest has recently received a new impetus due to research on self adaptive systems. The purpose of using performance indices in self adaptive systems is the same as for previous work; namely, to determine the optimum values of system parameters which may be varied to optimize system performance. The unique factor in an adaptive system is that the system itself performs this optimization.

The Performance Index replaces the usual design specifications for a system, i.e., instead of specifying that a system have a certain bandwidth, rise time, etc., it is only necessary to specify that the system have a certain (usually minimum) value of performance index.

This study was undertaken to determine whether or not Performance Indices should be used to evaluate and specify control systems. Performance Indices are judged here on the basis of their ability to select systems with good overall transient response when such factors as rise time, overshoot, and settling time are considered. Thus, if a performance index is rejected, this is not meant to imply that it is not valuable or acceptable for specific applications.

Almost all of the performance indices considered are based on step inputs. See the Final Report, Volume 1 [1], for the ramifications of using step inputs for system evaluation.

It is not possible to say that one performance index, such as ITAE or ITSE, is the best because requirements vary. One index may be more

applicable in certain applications than the others. Thus, it is desirable
to have a table of indices accompanied by data from which a designer can
choose the index most applicable to his requirements.

There are a few general rules which can be followed in the selection
of a general Performance Index. However, the relative weighting of these
factors is difficult to determine in general. These rules are an
elaboration of comments by Graham and Lathrop [53].

1. A general performance index should lead to systems of higher
   order, as well as second order, which judgment indicates are
   good systems when their overall response is considered. This
   property is called reliability.

2. A performance index should be selective. That is, the optimum
   value of system parameters should be clearly discernable from
   some characteristic, such as minimum, zero, or maximum value
   of a plot of the performance index value versus system para-
   meters.

3. The ease with which a performance index can be applied is a
   consideration.

The following Performance Indices are considered in detail in this
study:

IRAR (Impulse Response Area Ratio)

Logarithmic Decrement

Control Area

Weighted Control Area

IAE (Integral Absolute Value of Error)

ISE (Integral Squared Error)

rms Error

Solution Time

Fett's Criterion

Static Error Bandwidth Ratio

Gain Bandwidth Product

Beta

Bellman's Performance Index

GEF (Generalized Error Function)

Glover's Performance Index

Zaborszky and Diesel's Index

ITAE (Integral of Time Multiplied by Absolute Error)

ITSE (Integral of Time Multiplied by Squared Error)

ISTSE (Integral of Squared Time Multiplied by Squared Error)

ISTAE (Integral of Squared Time Multiplied by Absolute Value
of Error)

Aizerman's Performance Index

Rekasius's Performance Index

It was the original intent of this work to investigate the use of performance indices as an important factor for general system evaluation. It is now clear that this is impractical from the Air Force's point of view at the present time. With the current state of the art performance indices can be used only for system design, and perhaps as an aid to engineering judgment in the evaluation situation. It is hoped that in the future the confidence that comes with extensive use will make possible the application of a performance index as a major factor in the acceptance or rejection of a control system. System specifications that were recommended in Vol. 1 [1] have been in use for many years, while most performance indices are relatively new and still in the research stage. Thus they fail to pass the test of familiarity and wide experience. It is

hoped that the Air Force and industry will continue research in the performance index area, since it is obvious that this approach to system specification is more general and versatile than that presently recommended.

Even though it is outside the scope of this project to make recommendations in the design area, it is desirable to report developments which have occurred during the course of this research. With this preface the following recommendations can be made. The performance indices considered to be among the best of those presently in use for the general synthesis of systems are: ITAE, ISTSE, ITSE, and ISTAE. Data are shown for all these indices except ISTAE to support the position stated. The inclusion of ISTAE is justified because of its similarity to ITAE, although it places more emphasis on speed of response than ITAE. It is realized that one might want to use Aizerman's [54] method or consider such performance indices as the suggested by Bellman [52]. However, the additional study required to make definite statements on their applicability to design will not be undertaken, since it is outside the scope of the work at hand, and this additional work would not affect the recommendation on system evaluation. Both of these last named methods appear interesting at present evaluation, and it is important that further research be undertaken to prove or disprove their utility.

The data presented for ITAE, ITSE, and ISTSE are limited in quantity, but they should prove adequate for the selection of one of the indices over the others. All of the data are based on systems of the unity numerator type. This allows a more direct comparison of the indices. Of course, it would be essential to have sets of data for other types of systems if these IP are to be used in practice.

When a system is optimized with a Performance Index, the parameters
of the system are usually adjusted until a minimum value of the Perform-
ance Index is obtained. Unfortunately, there is no assurance that only
one minimum point exists. If the optimization is performed on a computer,
one naturally wonders if another minimum point exists which has a smaller
value of the Performance Index. The logical approach to answer this
question is the mathematical solution of the problem. The problem of
finding the number of minimum points, their value, and the rate of change
of Performance Index near the optimum points presents no formal mathe-
matical difficulties. However, the labor involved in obtaining numerical
answers for even a fourth order system is formidable if a digital computer
is not used. Even the problem of preparing the algebraic equations re-
quired before a computer program can be started for higher order systems
becomes very time consuming, and this is after the Performance Index has
been obtained in terms of system parameters. It is practical to solve
specific problems, such as designing a particular system, but the work
required to study a whole class of systems such as unity numerator systems
through the eighth and on is too large to perform on the present project.

For completeness, this report includes some methods available in
the literature for obtaining the mathematical solutions of Performance
Indices.

In conclusion, it is emphasized that this is a research area in
which it would be premature to make recommendations for systems evaluation.
Work is currently being done in this area and it enjoys the attention of
the professional societies. However, much more work remains to be done.
It is believed that further extension of Aizerman's work would be especially
fruitful.

4.2 Discussion of Available Performance Indices

4.2.1 IRAR (Impulse Response Area Ratio)

Abstract and Conclusions

IRAR is derived for a second order system in terms of the system damping ratio, $\mathcal{J}$. The IRAR has been obtained for several systems and used to define an equivalent $\mathcal{J}$ by analogy to a second order system. The percent overshoot of the system is compared to the percent overshoot of second order system on the basis of equivalent $\mathcal{J}$ being equal to the actual $\mathcal{J}$ of the second order system. The results lead to the conclusion that knowing the IRAR of a general system does not directly indicate commonly used system characteristics. It is not necessary to convert the IRAR data to an equivalent $\mathcal{J}$, as is done in this study. The systems could have been compared to a second order system directly by using IRAR. However, the IRAR is related to an equivalent damping ratio in this study because of the mathematical relationship between IRAR and $\mathcal{J}$ for second order systems, and to determine if $\mathcal{J}$ could be extended to higher order systems by using IRAR. The results show that $\mathcal{J}$ cannot be extended using IRAR.

IRAR is not recommended for use as a general performance index.

Discussion

IRAR is a measure of the relative stability of a system. It can be determined mathematically as a function of the damping ratio, $\mathcal{J}$, for a second order system, or determined from response data for any system. It is defined as the negative of the positive area under the impulse response curve divided by the negative area under the impulse response curve i.e.

$$\text{IRAR} = -\frac{A^+}{A^-}$$

(4.1)

IRAR is derived in the following for a second order system with a closed loop transfer function

$$\frac{C}{R} = \frac{\omega_n^2}{s^2 + 2\gamma\omega_n s + \omega_n^2} \qquad (4.2)$$

The weighting function of the system if obtained by taking the inverse Laplace transform of equation (4.2)

$$W(t) = \frac{\omega_n}{\sqrt{1-\gamma^2}} e^{-\gamma\omega_n t} \sin\omega_n\sqrt{1-\gamma^2}\, t \qquad (4.3)$$

The first cycle of the response is completed when

$$\omega_n\sqrt{1-\gamma^2}\, t = 2\pi$$

$$t = \frac{2}{\omega_n\sqrt{1-\gamma^2}}$$

The "n"th cycle of response is completed when

$$t = \frac{2\pi n}{\omega_n\sqrt{1-\gamma^2}}$$

The positive area under the impulse response curve of any cycle is now obtained by integration.

$$A_{+n} = \frac{n}{\sqrt{1-\gamma^2}} \int_{\frac{2\pi n}{\omega_n\sqrt{1-\gamma^2}}}^{\frac{\pi}{\omega_n\sqrt{1-\gamma^2}}(2n+1)} e^{-\gamma\omega_n t}\sin\omega_n\sqrt{1-\gamma^2}\, t\, dt$$

$$= \frac{\omega_n}{\sqrt{1-\gamma^2}}\left[e^{-\gamma\omega_n t}\, \frac{-\gamma\omega_n\sin\omega_n\sqrt{1-\gamma^2}t \ -\omega_n\sqrt{1-\gamma^2}\cos\omega_n\sqrt{1-\gamma^2}t}{\gamma^2\omega_n^2 \ \omega_n^2(1-\gamma^2)} \right]_{\frac{2\pi n}{\omega_n\sqrt{1-\gamma^2}}}^{\frac{\pi}{\omega_n\sqrt{1-\gamma^2}}(2n+1)}$$

$$A_{+n} = \frac{\omega_n}{\sqrt{1-\zeta^2}} \left[e^{\dfrac{-\pi \zeta(1+2n)}{\sqrt{1-\zeta^2}}} \omega_n \sqrt{1-\zeta^2} + \omega_n \sqrt{1-\zeta^2} \; e^{\dfrac{-2\pi n \zeta}{\sqrt{1-\zeta^2}}} \right]$$

$$= e^{\dfrac{-\pi \zeta(2n+1)}{\sqrt{1-\zeta^2}}} + e^{\dfrac{-2\pi n \zeta}{\sqrt{1-\zeta^2}}}$$

$$A_{+n} = \sum_{n=0}^{\infty} \left[e^{\dfrac{-\pi \zeta(2n+1)}{\sqrt{1-\zeta^2}}} + e^{\dfrac{-2\pi n \zeta}{\sqrt{1-\zeta^2}}} \right] \tag{4.4}$$

The negative area under the impulse response curve of any cycle is now

obtained by integration.

$$A_{-n} = \frac{\omega_n}{\sqrt{1-\zeta^2}} \int_{\dfrac{\pi(2n+1)}{\omega_n\sqrt{1-\zeta^2}}}^{\dfrac{2\pi(n+1)}{\omega_n\sqrt{1-\zeta^2}}} e^{-\zeta \omega_n t} \sin \omega_n \sqrt{1-\zeta^2}\, t\, dt$$

$$= \frac{\omega_n}{\sqrt{1-\zeta^2}} \left[e^{-\zeta \omega_n t} \frac{(-\zeta \omega_n \sin \omega_n \sqrt{1-\zeta^2}\, t - \omega_n \sqrt{1-\zeta^2} \cos \omega_n \sqrt{1-\zeta^2}\, t)}{\zeta^2 \omega_n^2 + \omega_n^2 (1-\zeta^2)} \Bigg|_{\dfrac{\pi(2n+1)}{\omega_n\sqrt{1-\zeta^2}}}^{\dfrac{2\pi(n+1)}{\omega_n \sqrt{1-\zeta^2}}} \right.$$

$$= - e^{\dfrac{-\zeta 2\pi(n+1)}{\sqrt{1-\zeta^2}}} - e^{\dfrac{-\zeta \pi(2n+1)}{\sqrt{1-\zeta^2}}}$$

$$A_{-n} = -\sum_{n=0}^{\infty} \left[e^{\dfrac{-\zeta 2\pi(n+1)}{\sqrt{1-\zeta^2}}} + e^{\dfrac{-\zeta \pi(2n+1)}{\sqrt{1-\zeta^2}}} \right] \tag{4.5}$$

The IRAR is now obtained by applying the definition of Equation (4.1)

$$IRAR = -\cfrac{\displaystyle\sum_{n=0}^{\infty}\left[e^{-\frac{\zeta\pi(2n+1)}{\sqrt{1-\zeta^{2}}}} + e^{-\frac{2\pi n\zeta}{\sqrt{1-\zeta^{2}}}} \right]}{\displaystyle\sum_{n=0}^{\infty}\left[e^{-\frac{2\zeta\pi(n+1)}{\sqrt{1-\zeta^{2}}}} + e^{-\frac{\zeta\pi(2n+1)}{\sqrt{1-\zeta^{2}}}} \right]}$$

$$IRAR = -\cfrac{\displaystyle\sum_{n=0}^{\infty} e^{-\frac{2\pi\zeta}{\sqrt{1-\zeta^{2}}}}\left[1 + e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^{2}}}} \right]}{\displaystyle\sum_{m=0}^{\infty} e^{-\frac{2\pi n\zeta}{\sqrt{1-\zeta^{2}}}}\, e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^{2}}}}\left[1+e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^{2}}}} \right]}$$

$$IRAR = e^{\frac{\pi\zeta}{\sqrt{1-\zeta^{2}}}} \qquad\qquad (4.6)$$

IRAR is plotted in Fig. 4.1 as a function of ζ for the second order system of Eq. (4.2). For small values of ζ the criterion is insensitive, i.e., a relatively large change in ζ yields a small change in IRAR with the result that knowing IRAR does not give an intuitive notion of the relative stability. For large values of IRAR the opposite is true. Then the criterion is extremely sensitive, a small change in ζ results in a large change in IRAR. The exponential nature of Eq. (4.6) suggests using a logarithmic plot; however, the sensitivity appears as bad or more inadequate on a logarithmic plot. The sensitivity is further illustrated in Fig. 4.2 by plotting ζ versus the rate of change of IRAR with ζ .

IRAR has been calculated from computer data for the optimum ITAE unity numerator systems through the eighth order. It has been obtained by graphical methods from the step function response data given in refer-
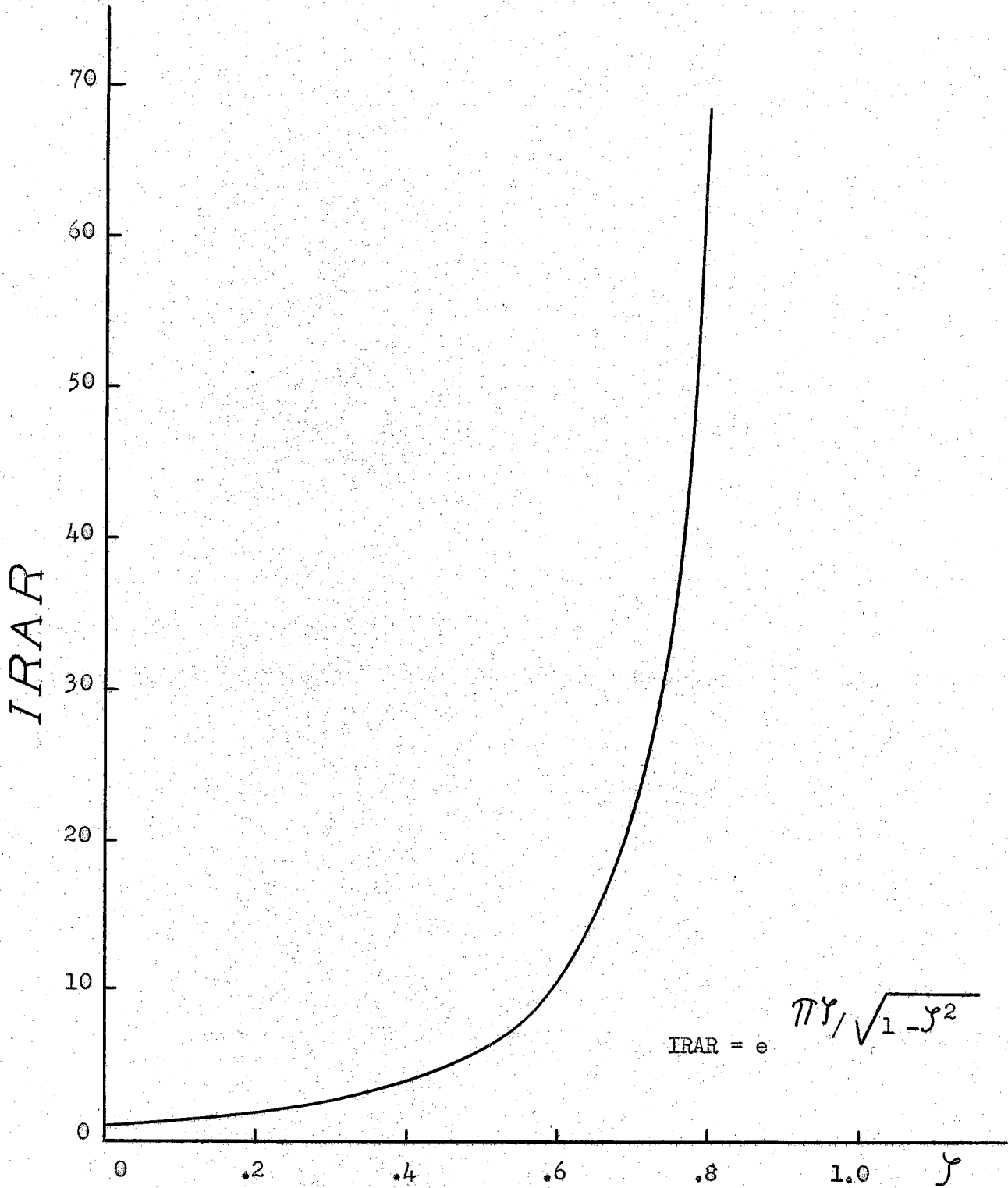
$$IRAR = e^{\pi \gamma / \sqrt{1 - \gamma^2}}$$

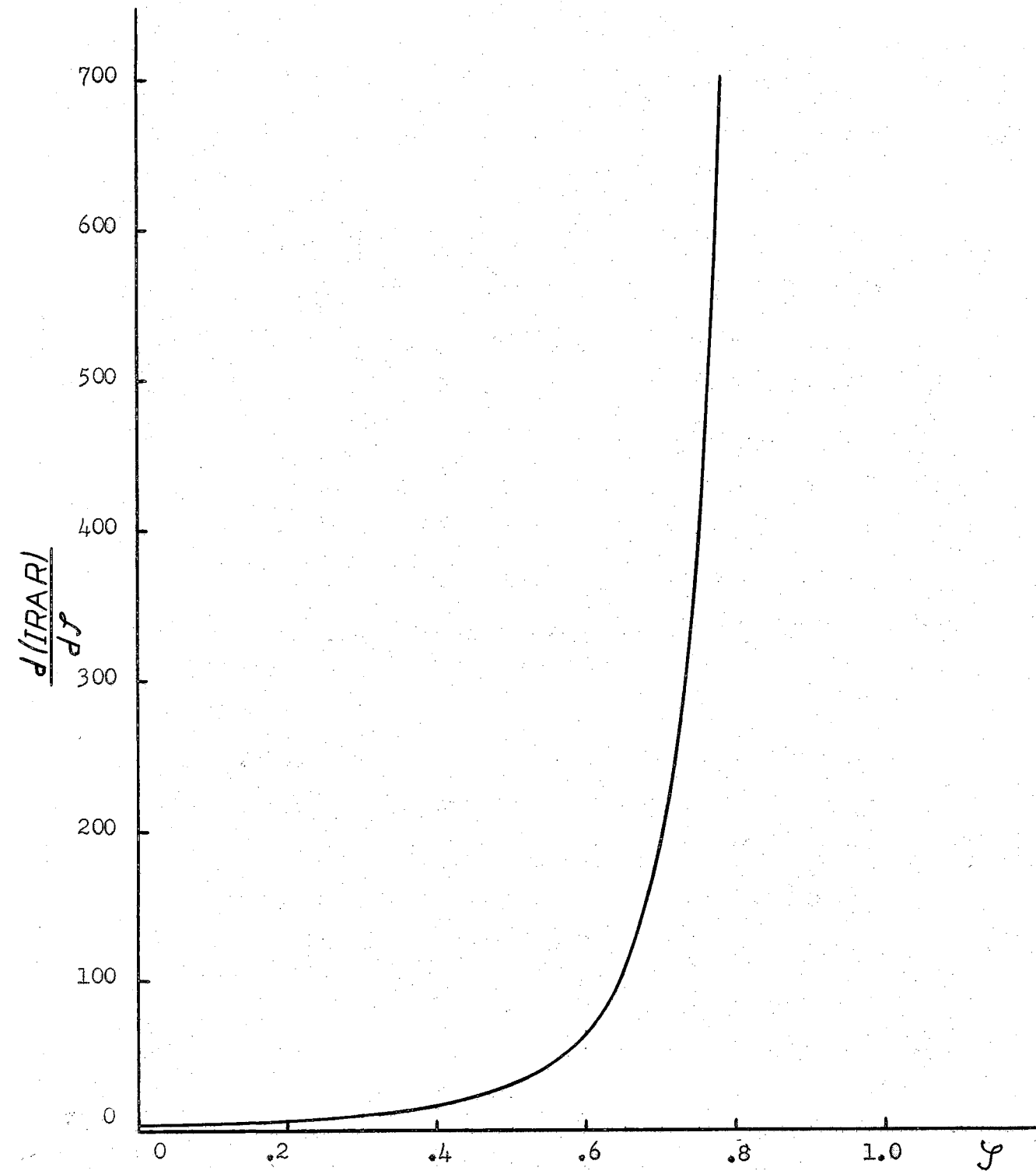Figure 4-1

IRAR vs. γ - Second Order System

Figure 4-2

$\dfrac{d(IRAR)}{d\gamma}$ vs. γ for a Second Order System

ence [55] for the zero-velocity-error systems through the sixth order. Although $\mathcal{J}$ is defined only for a second order system, the value of IRAR for each system was used to obtain an equivalent $\mathcal{J}$ by using Fig. 4.1. This equivalent $\mathcal{J}$ is plotted in Fig. 4.3 as a function of percent overshoot for each system where the number by each point represents the order of the system. The solid curve is exact data for a second order system defined by Eq. (4.2). The purpose of the data is to determine if, in general, IRAR leads to an equivalent $\mathcal{J}$ which is related to percent overshoot identical to the relationship existing between $\mathcal{J}$ and percent overshoot in the second order system of Eq. (4.2). Fig. 4.3 shows that a correlation does exist. Fig. 4.4 was derived from Fig. 4.3 to determine the accuracy of the criterion in predicting percent overshoot. The percent overshoot error was determined by the formula:

$$\% \text{ error} = \frac{\% \text{overshoot of general system} - \% \text{ overshoot of second order system}}{\% \text{ overshoot of second order system}}$$

(4.7)

Fig. 4.4 shows that the accuracy of using IRAR to predict the percent overshoot of a general system is inadequate. While other quantities such as settling time, which are a measure of relative stability, could be examined to determine if better correlation exists between the IRAR of a second order system and a general system for determining system characteristics, it is believed that the work would not be fruitful in view of the results that have been obtained.
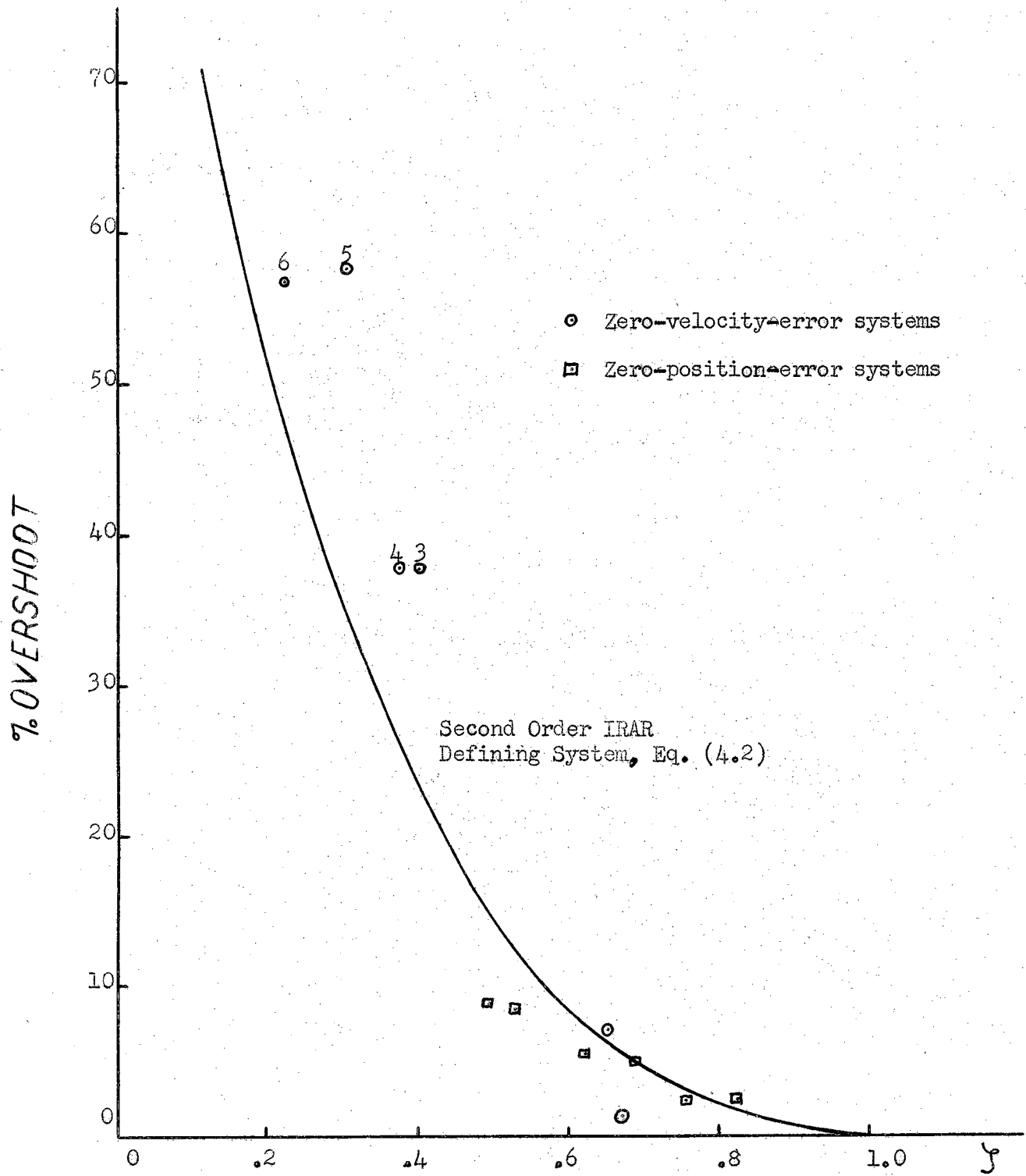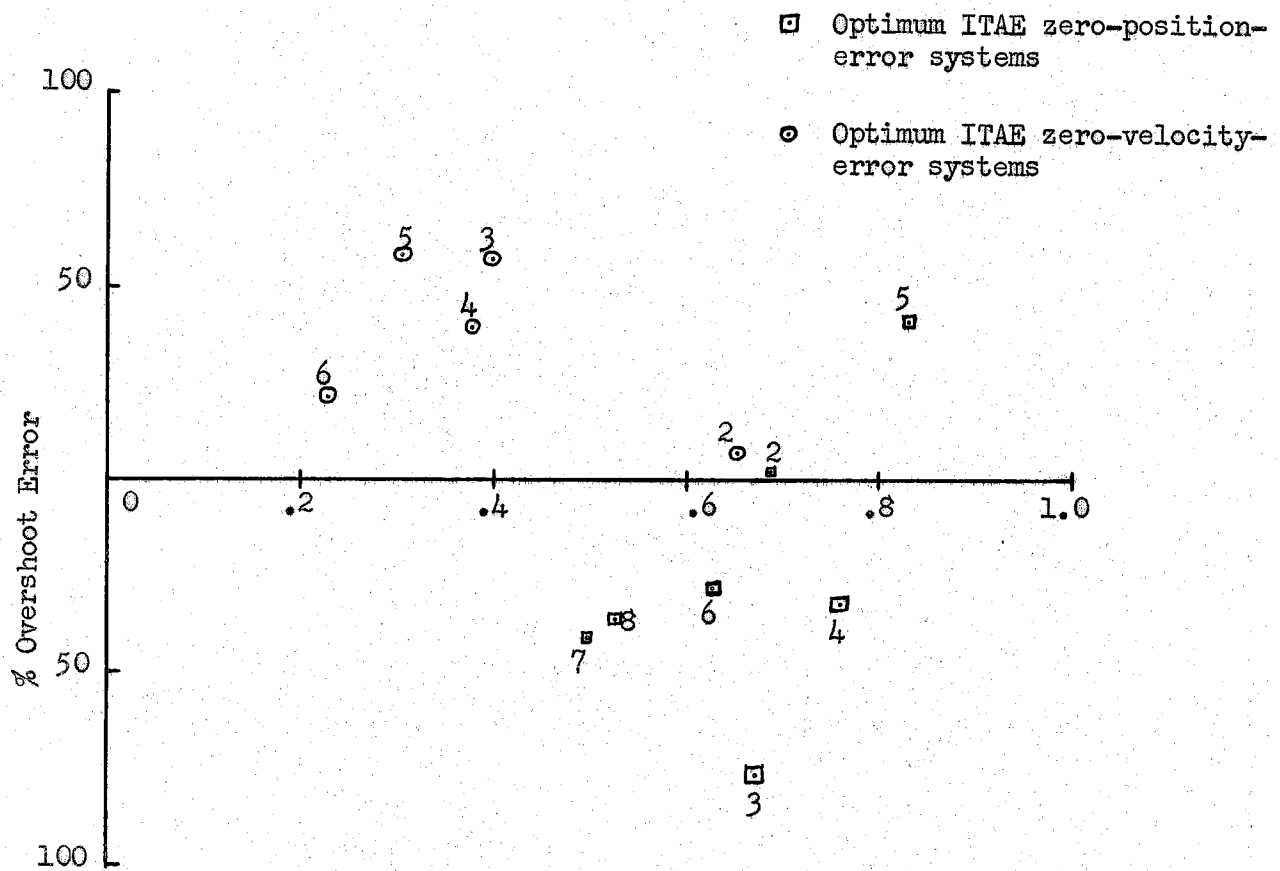
Figure 4-3

% Overshoot vs. Equivalent γ

Equivalent γ

Figure 4-4

Overshoot Error for Several Systems as a Function of
Equivalent γ with a Second Order System Used as a
Reference

Optimum ITAE Zero-Position-Error Systems Data
From Computer Study

System Order	IRAR	% Overshoot	Equivalent ζ	% Overshoot Eq.(4.2)	% Overshoot error
2	19.4	4.9	.688	4.8	+ 2.1
3	17.6	1.3	.671	5.7	- 77.2
4	37.2	2.13	.76	3.1	- 31.3
5	108.4	2.39	.83	1.7	+ 40.6
6	12.8	5.46	.628	7.6	- 28.2
7	6.5	8.94	.5	15.2	- 41.2
8	7.44	8.16	.532	12.7	- 35.8

Optimum ITAE Zero-Velocity-Error Systems
Data Obtained from Ref. [53], p. 283.

System Order	IRAR	% Overshoot	Equivalent ζ	% Overshoot Second Order System, Eq.(4.2)	% Overshoot error
2	15.5	7	.655	6.3	+ 6.35
3	4.08	37.9	.405	24	+ 57.8
4	3.87	37.9	.38	27.1	+ 39.8
5	2.61	55.5	.31	35	+ 58.7
6	2.0	56.8	.23	46.5	+ 22.2

Both sets of data are the average of two independent sets of graphical calculations. The third significant figure of the above data is not justified by the accuracy of calculations.

IRAR Data for Fig. 4.3 and Fig. 4.4

Table 4.1

4.2.2 Logarithmic Decrement

Abstract and Conclusions

Logarithmic decrement is a measure of relative stability. Its usefulness is limited to second order systems. For a second order system it is equal to twice the logarithm of IRAR.

Logarithmic decrement is not recommended for use as a general performance index.

Discussion

Logarithmic decrement is defined as the natural logarithm of the ratio of the maximum response overshoot during oscillation to the slightly smaller maximum response overshoot one cycle later (Skilling [56], p. 108).

The logarithmic decrement for a second order system such as the one defined by Eq. (4.2) is

$$L. D. = \frac{2\pi \varsigma}{\sqrt{1-\varsigma^2}} \qquad (4.8)$$

Reference to Eq. (4.6) shows that logarithmic decrement is equal to twice the logarithm of IRAR, i.e.,

$$L. D. = 2 \log IRAR \qquad (4.9)$$

IRAR was found to be inadequate as a performance index for systems of order higher than two. IRAR is based on the area under the impulse response curve. Logarithmic decrement takes into account only the amplitude of the response curve and thus is more sensitive to the shape of the response curve than IRAR. The inadequacy of extending logarithmic decrement to general systems can be illustrated by considering a practical third order system with the weighting function

$$W(t) = K_1 e^{-\beta t} + K_2 e^{-\alpha t} \cos (\omega t + \theta). \qquad (4.10)$$

- 63 -

For $\beta \gg \alpha$ it is easily seen that the logarithmic decrements for any two cycles of response are not equal. Logarithmic decrement has limited usefulness as a performance index.

4.2.3 Control Area

Abstract and Conclusions

The value of control area has been determined analytically for a second order system with a unit step input to be:

$$\text{Control Area} = 2\mathcal{F} \qquad (4.11)$$

The extremal values of this criterion are of no benefit in determining whether or not a system is of value. The only way this criterion could be used would be by analogy to some standard system, such as a second order system; however, it was shown for IRAR that this leads to erroneous results.

Control area is not recommended for use as a general performance index.

Discussion

Oldenbourg, Sartorius ([57], p. 66) and Nims ([58], p. 606) have suggested the control area criterion based on the minimization of the integral

$$\text{Control Area} = \int_0^\infty e\,dt \qquad (4.12)$$

for zero-displacement-error systems with a step function input. This integral gives the difference of the positive and negative area under the error versus time curve.

The analysis that follows shows the behavior of this criterion for a second order system with the weighting function:

$$W(t) = \frac{e^{-\mathcal{F}t}}{\sqrt{1-\mathcal{F}^2}} \sin \sqrt{1-\mathcal{F}^2}\; t. \qquad (4.13)$$

The output for a step of input with amplitude A is obtained by convolution:

$$C(t) = \frac{A}{\sqrt{1-\mathcal{f}^2}} \left[\sqrt{1-\mathcal{f}^2} - e^{-\mathcal{f}t} \left(\mathcal{f} \sin \sqrt{1-\mathcal{f}^2}\, t - \sqrt{1-\mathcal{f}^2} \cos \sqrt{1-\mathcal{f}^2}\, t \right) \right]$$

(4.14)

The error can now be obtained by substracting the output from the input:

$$e(t) = R(T) - C(t) = \frac{A\, e^{-\mathcal{f}t}}{\sqrt{1-\mathcal{f}^2}} \left(\mathcal{f} \sin \sqrt{1-\mathcal{f}^2}\, t + \sqrt{1-\mathcal{f}^2} \cos \sqrt{1-\mathcal{f}^2}\, t \right.$$

(4.15)

The control area may now be determined by integrating Eq. (4.15)

$$\int_0^\infty e\, dt = \frac{A}{\sqrt{1-\mathcal{f}^2}} \left\{ \mathcal{f}\, e^{-\mathcal{f}t} \left(-\mathcal{f} \sin \sqrt{1-\mathcal{f}^2}\, t - \sqrt{1-\mathcal{f}^2} \cos \sqrt{1-\mathcal{f}^2}\, t \right) \right.$$
$$\left. + \sqrt{1-\mathcal{f}^2}\, e^{-\mathcal{f}t} \left(-\mathcal{f} \cos \sqrt{1-\mathcal{f}^2}\, t + \sqrt{1-\mathcal{f}^2} \sin \sqrt{1-\mathcal{f}^2}\, t \right) \right\} \Big|_0^\infty$$

$$\text{Constrol Area} = 2\,\mathcal{f}\,A$$

(4.16)

Eq. (4.16) is plotted in Fig. 4.5 with A equal to unity. Control Area varies from zero for a system with a damping ratio of zero to infinity for a system with an infinite damping ratio. The extremal values of this criteria for a second order system in no way indicate an optimum system, i.e. the criterion has no selectivity. For a second order system (or any completely defined system) curves of control area versus any desired system characteristic can be plotted so that control area can be assigned a meaning in terms of common concepts; however, this is precisely what one wants to avoid for a general performance index, unless knowing the relationship of the criterion to a system characteristic for a particular system leads to knowledge of this characteristic of systems in general. There is nothing unique about control area which would indicate that this is true.
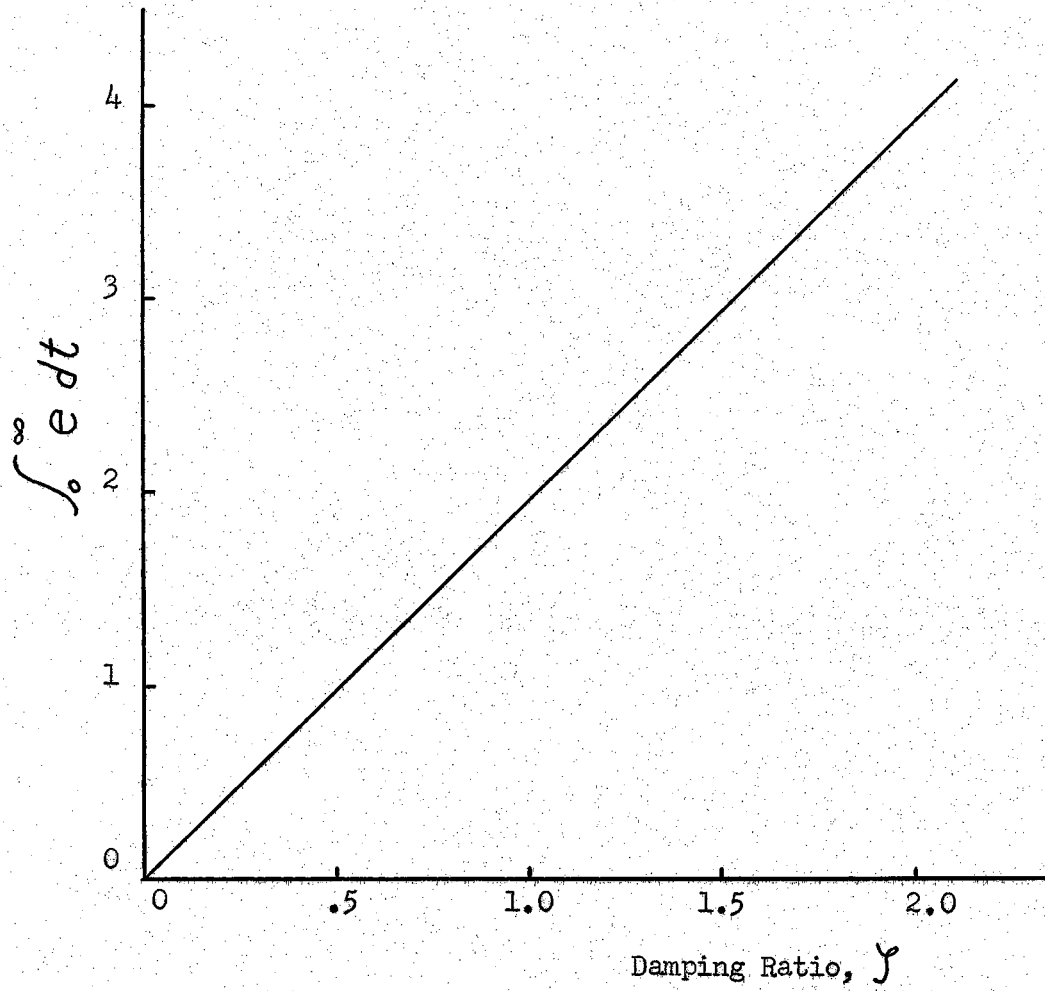
Damping Ratio, γ

Figure 4-5

Control Area vs. γ for $\frac{C}{R} = \dfrac{1}{s^2 + 2\gamma s + 1}$

4.2.4 Weighted Control Area

Abstract and Conclusions

This criterion has been studied analytically and on an analog computer. It has been shown that it yields an optimum third order system which is unstable.

The weighted control area performance index is not recommended as a general performance index.

Discussion

Nims ([58], p. 606) has suggested that the control area criterion could be modified by time weighting the error as shown in Eq. (4.17)

$$\text{Weighted Control Area} = \int_0^\infty t \, e(t) \, dt \qquad (4.17)$$

Weighted control area as a function of damping ratio is plotted in Fig. 4.6 for a second order system with a step input. This graph is reproduced from reference ([53], p. 276).

The extremal values of this criterion are of no value in selecting a good system, but the zero value of the criterion selects a second order system with a damping ratio equal to 0.5, which is usually considered satisfactory. This suggests the possibility of using the minimum value of the absolute magnitude of the criterion as a figure of merit. To this end a third order system

$$\frac{C(s)}{R(s)} = \frac{1}{s^3 + bs^2 + fs + 1}$$

was studied. Knothe (in an unpublished work) of AFMDC showed analytically that the system has a zero value of performance index when
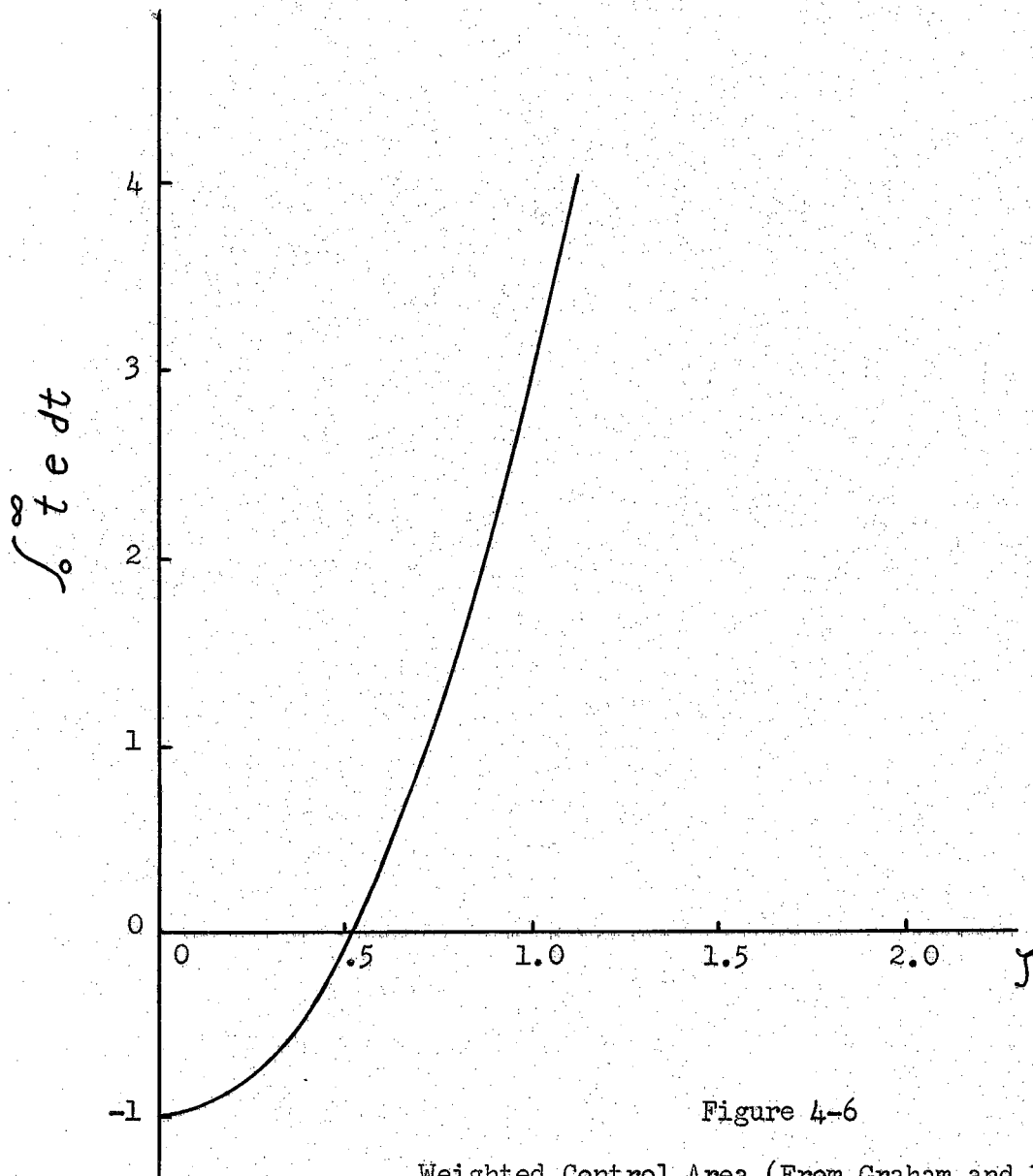
$$f^2 = b$$

- 67 -

Figure 4-6

Weighted Control Area (From Graham and Lathrop)

$$\text{for } \frac{C}{R} = \frac{1}{s^2 + 2\gamma s + 1}$$

This relation was verified by Haldeman of AFMDC who showed that when $f = b = 1$ a transient will not be damped; i.e. the system is on the borderline between stability and instability even though it is optimum in the weighted control area sense.

This criterion is not recommended for use as a general performance index. Its lack of ability to select even a good third order system disqualifies it.

4.2.5 IAE (Integral of the Absolute Value of Error)

Abstract and Conclusions

IAE is applicable to second order systems but has inadequate selectivity for higher order systems.

IAE is not recommended for use as a general performance index.

Discussion

The IAE (integral of the absolute value of error) is defined by the equation:

$$IAE = \int_0^\infty |e(t)| \, dt \quad . \tag{4.18}$$

This criterion discriminates against total error independent of polarity, since the absolute value of error is used. A system is considered optimum in the IAE sense when it is adjusted to have minimum IAE to a step input. A method has been pointed out in the literature (Fiechesen [59], p. 244) for measuring this criterion with a standard rectifier type volmeter.

Graham and Lathrop ([53], p. 277) have found that the IAE criterion selects a second order system with a $\zeta = 0.7$. A step input was used in the study. The selectivity is adequate for a second order system, but their investigation of a third order system showed the selectivity to be inadequate. The criterion value as a function of system parameters is

shown in Fig. 4.8A. This picture shows that there is no change in the criterion value when the parameter b is varied from 1.25 to 2.0 and only a 10% change in IAE for a 2 to 1 change in parameter c. Fig. 4.8C shows the variation of the output as a function of these system parameters. Graham and Lathrop report that IAE is even less selective for higher order systems.

The inability of this criterion to make a definite selection of good high order systems disqualifies it for general use.

4.2.6 ISE (Integral of Squared Error)

Abstract and Conclusions

ISE has been used primarily because of mathematical convenience. It selects systems which are underdamped. The selectivity is also inadequate.

ISE is not recommended for use as a general performance index.

Discussion

The ISE (integral of squared error) criterion is defined by Eq. (4.19)

$$ISE = \int_0^\infty e^2(t)\, dt. \tag{4.19}$$

Although the criterion can be used with any input for which the integral converges, step inputs have been used in this discussion. This criterion discriminates against total error independent of polarity since error is squared. Hall [60] has shown that for a second order system ISE can be determined as a function of $\mathcal{S}$ and ω_o with a step input. The relationship is:

$$ISE = \frac{1 + 4\mathcal{S}^2}{4\mathcal{S}\omega_o} \tag{4.20}$$

This equation is plotted in Fig. 4.7 with ω_o equal to unity. The criterion has an optimum value when f is equal to .5 . A system is considered optimum when it is adjusted to have minimum ISE. . Hall [60] concluded that ISE selects systems which are too underdamped for many applications and that selectivity is poor.

Graham and Lathrop ([53], p. 277) have also found that the selectivity is poor. This is illustrated in Fig. 4.8A. Fig. 4.8C shows the output response as a function of system parameters, illustrating the results of inadequate selectivity. It can be seen in Fig. 4.8A that a change of parameter c from 1.6 to 2.4 results in only a 4.5% change in the criterion value.

ISE has been used primarily due to mathematical convenience. Using Parseval's theorem, frequency domain information can be used to evaluate ISE. However, the results obtained may be misleading. Newton, Gould, and Kaiser ([61], p. 46) work an example which leads to an unstable system for an optimum mathematical value of the criterion. The restrictions to impose on the mathematical solution are obvious in this case. Interpretation of the mathematical result in a higher order system may be formidable task.

Clark [62] has used ISE as a direct measure of the speed of response, percent overshoot, settling time, and other salient characteristics of the transient response. He defines error as being the difference between the system response and a desired response. The ISE criterion is applied in the same manner as others have used it but with the error as defined above. When ISE is very small, the system must be similar to the known model, hence, the characteristics of the system are known. It is recognized by Clark that the idea is useful for evaluation only when ISE is small.
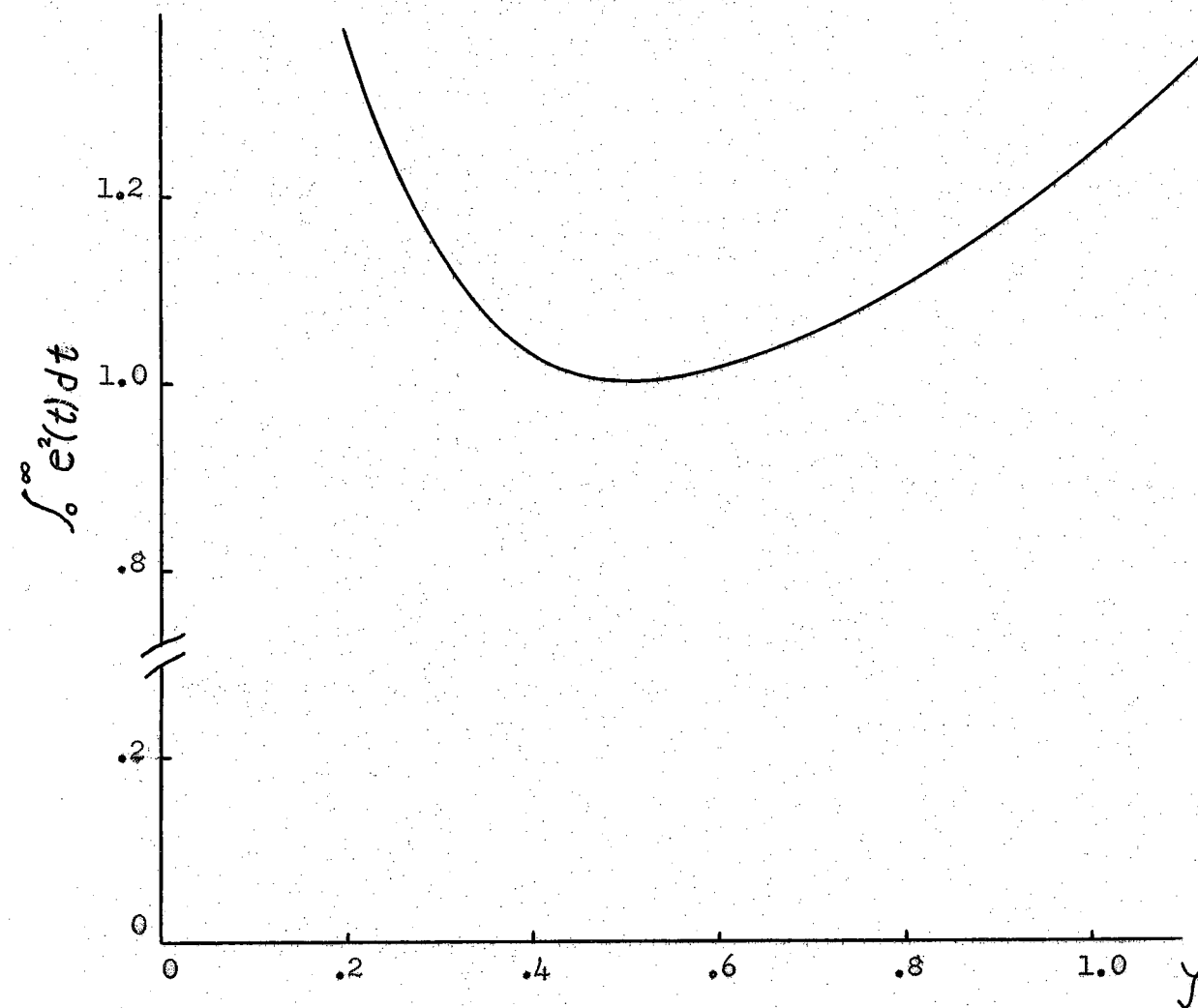
Figure 4-7

Integral of Squared Error vs. γ for a Second Order System

If ISE is small for the system under evaluation, (in the sense that Clark uses ISE) ISE is hardly needed, since the actual response could be superimposed on the desired response and an evaluation could be made. However, if ISE is not small, the method fails, according to the author. Therefore, this idea must be rejected for system evaluation in the sense desired in this work.

For self adaptive systems, if the system can be made to respond approximately like the model (i.e. the desired response), the use of ISE would insure proper performance, and hence be good for system evaluation or model identification. However, it appears likely that other criteria could be used which are easier to instrument such as IAE. For analysis, it is advantageous to use ISE, but for building hardware, the case where it is agreed that ISE performance is acceptable, there is insufficient evidence to justify a recommendation in favor of ISE over other perform-ance criteria, e.g. IAE.

System synthesis is another distinct use of the method that should be considered. There is no reason why ISE can not be used in the same manner that Aizerman used a performance index; in fact, Clark has done so. Whether or not ISE is superior in a synthesis application of this type is not known.

It should be noted that Newton, et. al. [61] expressed the same philosophy as Clark when they said "the performance index is the integral-square value of the error between the ideal output and the actual output". They, of course, do not give detailed treatment to cases where the ideal output is a step function response, as Clark has done.

This criterion does not have sufficient merit to justify its recom-mendation for general use. In particular, its inability to select a good

higher order system disqualifies it.

4.2.7 rms Error

Abstract and Conclusions

This criterion has been used primarily due to mathematical convenience. Systems optimized by this criterion are unsatisfactory in many cases due to inadequate damping.

The rms error criterion is not recommended for use as a general performance index.

Discussion

The rms criterion is defined ([63],p. 309) by the equation,

$$\text{rms error} = \lim_{T \to \infty} \left[\frac{1}{2T} \int_{-T}^{T} e^2(t)\, dt \right]^{\frac{1}{2}} . \tag{4.21}$$

A large amount of literature ([63] , p. 309), [61] , (Truxal [64] , pp. 4-74), ([3] , p. 413) is available concerning this criterion, not because of its goodness, but primarily because of its mathematical convenience in systems concerned with stochastic inputs, although any input could be used for which the definition has meaning. Truxal ([64] , pp. 4-74) and ([3] , p. 413) points out that optimum rms error systems are not staisfactory in many cases. He says "A system may be comparatively unstable, being effective in rapidly reducing large errors but allowing undesirable long tails of error or excessive overshoot". The rms error criterion may be a good starting point but does not yield a good final system.

The reason this criterion selects a system with a relatively low degree of stability is that the error is squared and, hence, it weighs most heavily the large error and produces an optimum system which rapidly

reduces large error. The rapid reduction of large error results in large overshoots or low damping.

James, Nichols, and Phillips ([63] , p. 309) were motivated to use the rms criterion apparently because Wiener [65] used it for the analysis of stationary time series. They also point out that its wide usage is due to its mathematical convenience and because there is a highly developed body of mathematical knowledge built around mean square values. In their example on a radar automatic tracking system ([63] , p. 328) they point out that using the rms criterion led to almost the same results that were obtained using standard design techniques, plus some trial-and-error adjustments. In this example the criterion did not improve the design and, as already noted, the criterion can result in unsatisfactory systems. The fact that the final design is compared to the results from other methods may be good engineering, but this indicates a lack of confidence in the performance index.

This criterion is not recommended for use as a general performance index.

4.2.8 Solution Time

Abstract and Conclusions

This criterion chooses a good second order system. It chooses higher order systems which are underdamped.

Solution time is not recommended for use as a general performance index.

Discussion

This criterion has been defined in reference (Guillemin [66]) as follows:

"After a unit step function is applied, the time for the solution to

reach $\pm$ 5% of final value and not exceed it shall be a minimum for the 'optimum' transient response of systems of a given order." The criterion name is abbreviated as ST.

This criterion chooses a second order system with a $\mathcal{S}$ = 0.7, which is considered good. For higher order systems the response chosen as optimum becomes increasingly oscillatory and approaches neutral stability for a sixth order unity numerator zero-displacement-error system according to J. W. Froggatt, Jr. ([67] , p. 20). He also found that the criterion is not always precisely reproducable due to the nature of the criterion and its selectivity.

This criterion is unacceptable for systems above the fourth order.

4.2.9 Fett's Criteria

Abstract and Conclusions

The criteria have no meaning for an overdamped system. They have inadequate selectivity and are difficult to apply for an underdamped system.

Fett's criteria are not recommended for use as general performance indices.

Discussion

This criterion was suggested by G. H. Fett in the discussion of a paper by D. Graham and R. Lathrop ([53] , p. 287) on ITAE. He defined the criterion as being the value of the output displacement at the first overshoot, when a unit step displacement is applied to the input, multiplied by the time required to reach the maximum deflection. The criterion value then is a measure of the area on the displacement time curve of the rectangle bounded by the maximum deflection and the overshoot time. Due to the vagueness of the statement of the criterion, the meaning of

overshoot is not explicit. Froggatt considered four possible definitions of the criterion, all of which led to a second order system with a $\mathcal{f}$ = 0.5. He did not investigate systems higher than the third order because it was felt that at best the criteria would choose responses similar to the solution time criterion, which responses are unacceptable. The criterion has no meaning for a system without overshoot. For a third order system a small change in the optimum criterion value results in a large and irregular change in the nature of the response. Discontinuities and irregularities exist when the criterion value is plotted as a function of system parameters. This makes it difficult to determine the optimum value of the criterion.

4.2.10 Static Error Bandwidth Ratio

This performance index is discussed in the frequency domain specification section (Section 3.7) of Vol. 1 [1] . It is not recommended for use as a general performance index.

4.2.11 Gain Bandwidth Product

This performance index is discussed in the frequency domain specification section, Section 3.6 of Vol. 1 [1] . It is not recommended for use as a general performance index.

4.2.12 Beta

Abstract and Conclusions

Beta is the transfer function from the output to the input of a system; i.e., it is a function of the elements in the feedback path. This quantity is often designated as "H" in control systems. No reference has been found in the literature where beta is used in an electromechanical system other than those involving meter movements. It is used principally in feedback amplifiers. Beta used by itself has no meaning as a perform-

ance index. Beta multiplied by the system open loop gain is a performance index in that it is a measure of system error. The more general error constants contain the same information, hence, there is no justification for using beta.

Beta is not recommended for use as a general performance index.

Discussion

The earliest reference to beta found in the literature is by H. Black ([68] , p. 114) (1934) who defined beta as the "propagation of feedback circuit". Black used the quantity beta multiplied by the forward part of the system open loop gain. Nyquist ([69] , p. 126) used a product equal to this quantity but did not define beta. G. Happell and W. Hesselberth ([70] , p. 302) have defined beta as the voltage feedback to the input divided by the output voltage.

$$\beta = \frac{e_{fb}}{e_o}$$ (4.22)

This is identical to the quantity which is often denoted as "H" in conventional system block diagrams.

Beta is used extensively in the literature in conjunction with feedback amplifiers. No references have been found where this quantity is used with electro-mechanical systems other than systems involving meter movements. The product of system gain and beta is used in feedback amplifiers as a performance index because the feedback reduces distortion, effect of component variation, etc. This product is sometimes expressed in decibels and the number of decibels being fed back is used as a performance index. Since system gain is a function of frequency, the use of this performance index can lead to erroneous

conclusions, unless it is used only at the frequency at which the gain

beta product is specified. This is often done using the dc gain, in

which case the same information is available from the error constants

([1] , Section 4.8). The error constants are more general and are

recommended specifications.

4.2.13 ITAE (Integral of Time Multiplied by the Absolute Value of Error)

Abstract and Conclusions

This criterion has been treated extensively in the literature by

Graham and Lathrop ([53] , p. 273), ([71] , p. 10), ([72] , p. 153).

Only the essentials are repeated in this report. ITAE chooses good

unity numerator zero-position-error systems. The optimum ITAE zero-

velocity-error systems have excessive overshoot. ITAE is considered to

be one of the best performance indices available, but it is not recom-

mended for system evaluation (see the introduction for further discussion).

Discussion

Graham and Lathrop ([53] , p. 273) have suggested using a perform-

ance index defined by equation (4.23),

$$ITAE = \int_0^\infty t \left| e(t) \right| dt \ . \tag{4.23}$$

A system is optimized using a step of position input and is optimum when

ITAE is a minimum. In words, the criterion is called the integral of

time multiplied by the absolute value of error. ITAE evaluates system

error in a weighted manner which is intuitively good, in that it dis-

counts initial error, which is a basic limitation of all systems, and

magnifies error which persists in time. It discriminates against both

positive and negative error and evaluates all three of the important

quantities -- speed, stability, and accuracy.

ITAE can be evaluated by referring to data which is reproduced from reference [53] . Fig. 4.8B illustrates the superior selectivity of this criterion. The selectivity may be compared to two other criteria which are more selective than most other criteria, by referring to Fig. 4.8A. Although this data is for a third order system, Graham and Lathrop have found that the selectivity is good through the eighth order systems, the highest on which they reported. Fig. 4.9A shows the step function response of the optimum ITAE unity numerator systems through the eighth order. The criterion chooses a second order system with a f = 0.7. Fig. 4.9B shows the step function response of the optimum zero-velocity-error systems through the sixth order, and Fig. 4.9C shows the step function response of the optimum zero-acceleration-error systems from the third through the sixth order. The overshoot is excessive. This fact is sufficient to negate the possibility of using ITAE by itself to select systems of any type other then the zero-position-error type.

On the basis of selectivity and the ability to select good zero-position-error unity numerator systems, ITAE demonstrates that it is a superior performance index. It is also shown that ITAE does not lead to zero-velocity-and zero-acceleration-error systems which one would consider superior or even as good as those obtained by conventional design procedures. In addition, it is felt that results with unity numerator systems are not sufficient to insure good evaluation results with all non-unity numerator systems. This comment is applicable to all performance indices, and few have been studied as thoroughly as ITAE.
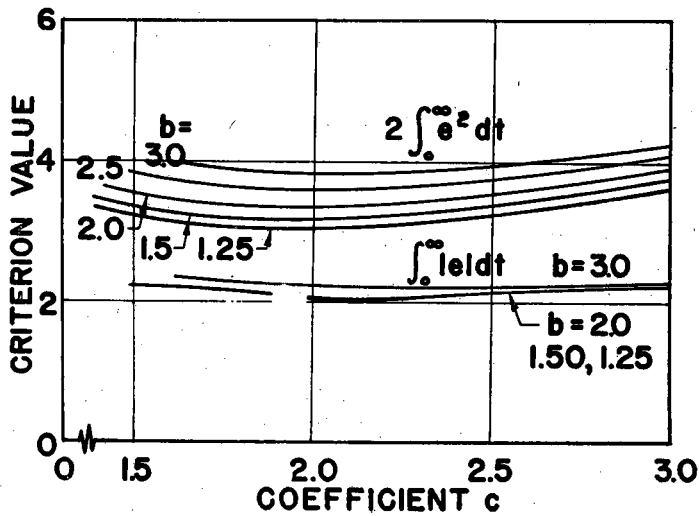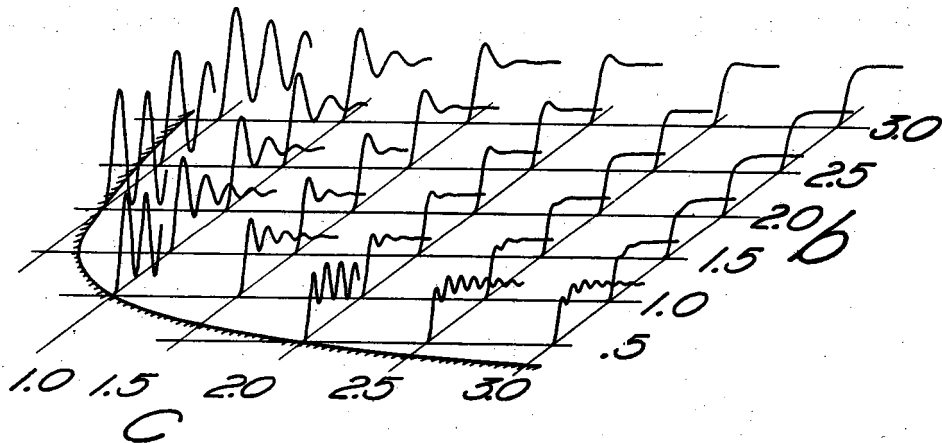
Integral of squared error and integral of absolute value of error criteria applied to the step-function responses of third-order systems

A
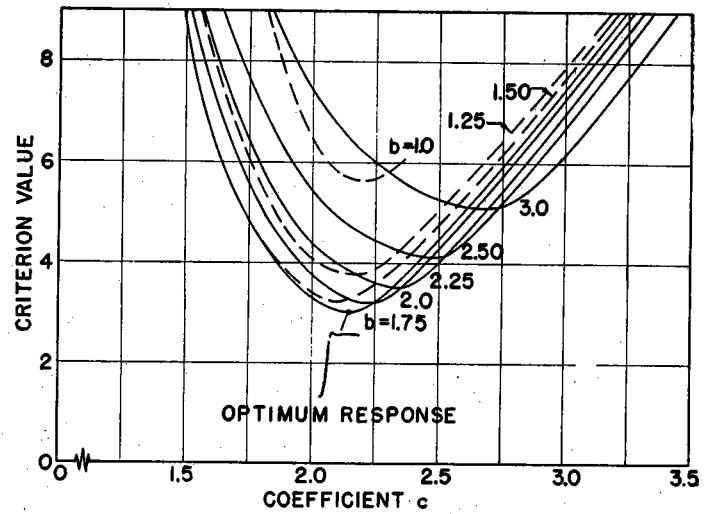
The integral of time-multiplied absolute value of error criterion applied to the step-function responses of third-order systems
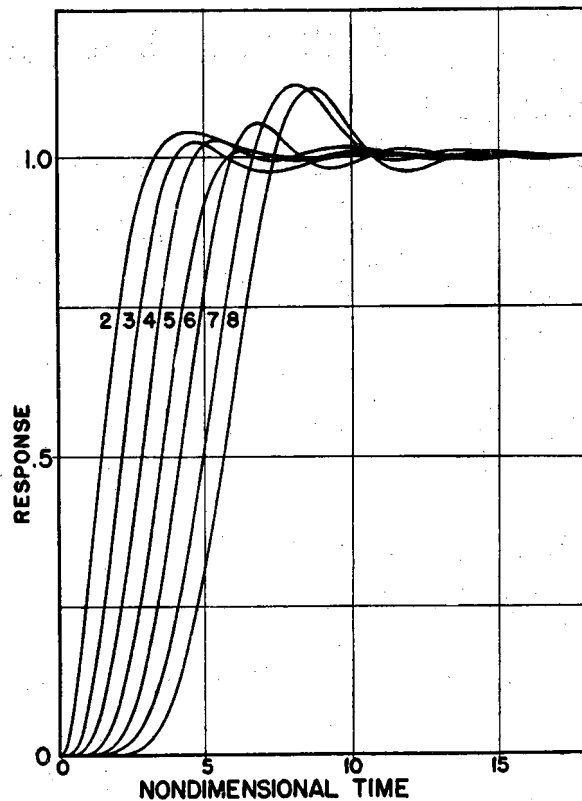
B

C

Step-function responses of third-order systems with the transfer function

$$\frac{C(s)}{R(s)} = \frac{1}{s^3 + bs^2 + cs + 1}$$

C

Fig. 4-8

Selectivity of Some Criteria
(From Graham & Lathrop)

Step-functions responses of the optimum unit-
numerator transfer systems, second to eighth orders. These
responses have a minimum integral of time-multiplied abso-
lute value of error

A
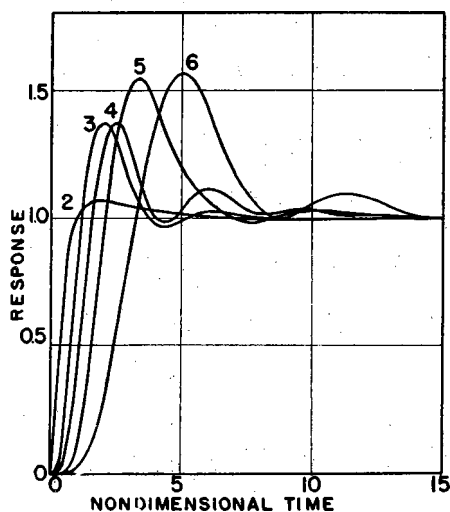
Step-function responses of the
optimum zero-velocity-error systems, second
to sixth orders

B

Fig. 26. Step-function responses of the opti-
mum zero-acceleration-error systems, third to
sixth orders

C

Fig. 4-9

Step Function Response of Optimum ITAE Systems (From Graham & Lathrop)

4.2.14 ITSE (Integral of Time Multiplied by Squared Error)
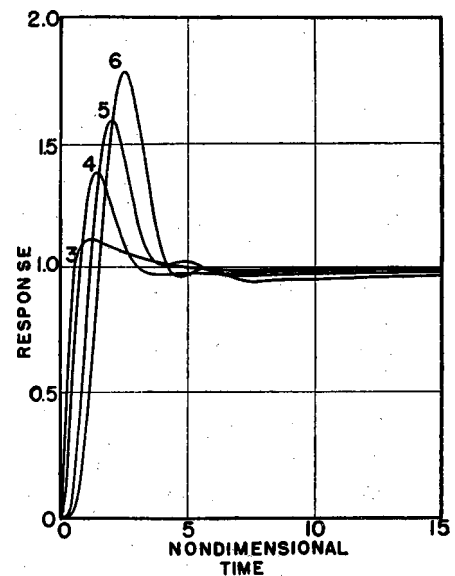
Abstract and Conclusions

ITSE is considered to be one of the best performance indices. However, it is not recommended to the Air Force for system evaluation (see the introduction for qualifications).

Discussion

ITSE is defined by the equation

$$ \text{ITSE} = \int_0^\infty t \ e^2(t)dt \tag{4.24}$$

A system is considered optimum when the above integral is a minimum. A step input is used for the evaluation.

This criterion weights time error in the same manner as ITAE, but weighs large error more than ITAE due to error being squared. ITSE proved to be satisfactory in two studies (Gibson [73]), (Gibson [74]) at Purdue University with adaptive systems.

ITSE has not received exhaustive study, but it has proved adequate in all cases where it has been used or studied. Data for unity numerator zero position error systems are shown in Fig. 4.10. This data is reproduced from work by (Stone [75]). From this data and from the results in reference [73] and [74] , it is concluded that ITSE is one of the best criteria considered in this report.

4.2.15 ISTSE (Integral of Squared Time Multiplied by Squared Value of Error)

Abstract and Conclusions

ISTSE is one of the best performance indices considered in this study. It has been thoroughly studied by Crow [76] for type one unity numerator systems. However, as indicated in the introduction, it is not recommended

Fig. 4-10

RESPONSE OF ITSE OPTIMUM UNITY NUMERATOR SYSTEMS

n—ORDER

2
3
4

OUTPUT

$\tau = \omega_0 t$

$$OUTPUT = 1 - \mathcal{L}^{-1} \frac{s^{n-1} + A_{n-1} s^{n-2} - - - - - + A_2 s + A_1}{s^n + A_{n-1} s^{n-1} + - - - - - - + A_1 s + 1}$$

-83-
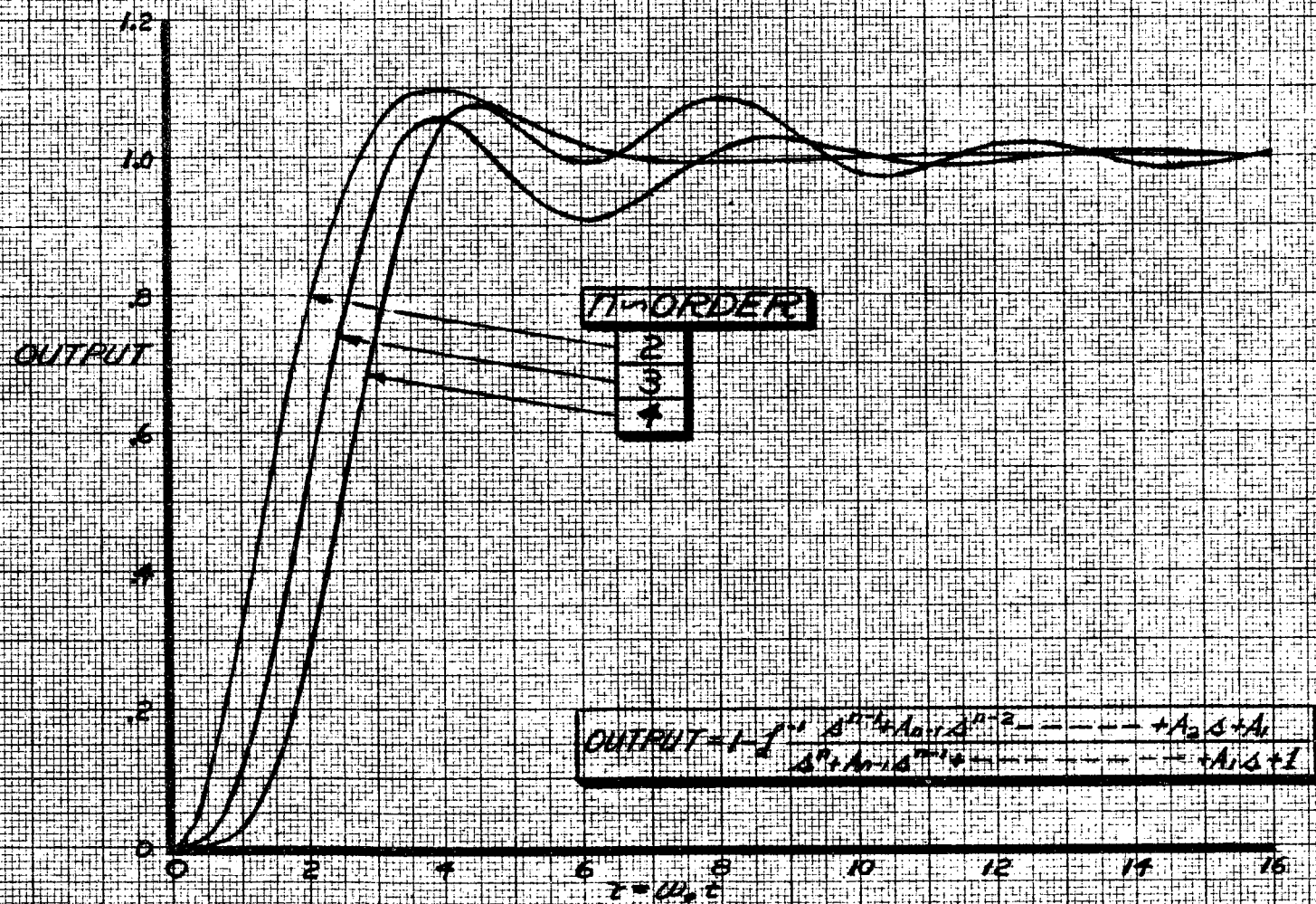
for system evaluation.

Discussion

This performance index is defined by the equation

$$ISTSE = \int_0^\infty t^2 e^2 (t)dt \qquad (4.25)$$

A system is considered optimum when the above integral is a minimum.
A step input is used for the evaluation.

Crow obtained optimum type one unitity numerator systems through
the eighth order by an analog computer study. These are not reproduced
here, see [76]. The step responses of the systems are judged to be good.
On the basis of this work ISTSE is considered to be one of the best per-
formance indices available. However, for system evaluation on specifica-
tion it can not be recommended for use by the Air Force. See the intro-
duction for details of this decision.

4.2.16 ISTAE (Integral of Squared Time Multiplied by Absolute Value of Error)

Abstract and Conclusions

ISTAE is considered to be a valuable performance index because of its
similarity to ITAE, ISTSE and ITSE. It has received little attention, but
it is believed that ISTAE is of value for those applications where minimiza-
tion of persistent error is especially important. For the reasons given in
the introduction it is not recommended for system evaluation.

Discussion

This performance index is defined by the equation

$$ISTAE = \int_0^\infty t^2 \; e(t) \; dt \qquad (4.26)$$

A system is considered optimum when the above integral is a minimum. A step input is used for the evaluation.

The salient feature of this performance index which makes it valuable is its ability to heavily penalize persistent error; i.e. the system must approach equilibrium rapidly. This statement is not based on data but by analogy to ITAE, ITSE and ISTSE.

4.2.17 Rideout and Schultz work

Abstract and Conclusions

Rideout and Schultz [77] have worked with performance criteria of the general form.

$$PI = \int_0^\infty F\left[e(t), t\right] dt \qquad (4.27)$$

They point out that criteria such as ISE, ITAE, etc. are special cases of equation (4.27). In their work, it is emphasized that criteria should not be chosen because they make the analysis problem easy. The availability of computers enables a designer to use complicated criteria. No new performance indices are presented which should be considered in this work; hence, no recommendations are necessary.

Discussion

Schultz and Rideout [77] have published one of the most comprehensive papers on performance indices. They divide the area on a historical basis and classify the work as past, present and future. The main value of the paper to a reader of this report is one organization of the material. The material was written with a different motivation than this report. It is a survey of the area, a classification of the different criteria, and an approach to performance criteria from an overall philo-

sphical view, while in this report it is desired primarily to examine the different criteria for their utility in system evaluation. Rideout and Schultz report on three papers [78] , [79] , [80] , not included in this report which are considered to be of importance to researchers working in this area. Their papers will not be discussed here because the ideas expressed are at the germinal stage and have not been developed sufficiently for the application of system evaluation.

4.2.18 Aizerman's Work and Its Extension

Abstract and Conclusions

This work is interesting and is a fruitful area for further research, however, it can not be used for system evaluation and specification at the present time by the Air Force.

Discussion

Aizerman's approach to system synthesis via performance indices is not philosophically different from what others have done, e.g. see Newton et. al., [61] . Aizerman uses a performance index to minimize the difference between system response and a desired response. For the desired response he uses a model which could be called a model performance index. To avoid confusion the performance index used to null the system and model will be referred to as the minimizing performance index.

A translation of Aizerman's [54] work is contained in Appendix B, so that only a brief description need be included here as a introduction to more recent work by Rekasius [81] .

Aizerman proposed the minimizing performance index

$$I = \int_0^\infty \left[e^2 + \tau_1^2 \dot{e}^2 + \tau_2^4 \ddot{e}^2 + \ldots + \tau_n^{2n} \left(\frac{d^n e}{dt^n} \right)^2 \right] dt \qquad i=1,2,\ldots,n \tag{4.2}$$

which he used to minimize the difference between the actual system response and the model, where

e is the error of the actual system

τ_i are constants available from the differential equation of the model.

It has been shown that when i=1 (the simplest case), the system can be made to approach a first order model with the characteristic equation

$$x + \tau_1^2 \dot{x} = 0 \qquad (4.29)$$

Unfortunately, the characteristic equation alone does not describe a system, so this result is misleading. It was also shown that the maximum deviation of the system from the model is

$$|\Delta x| \leqslant \left[\frac{I_{min} - I_{min\ min}}{\tau_1^2} \right]^{\frac{1}{2}} \qquad (4.30)$$

where

Δx is the maximum difference between the actual response x(t) of the optimum system and the model.

I_{min} is the value of the minimizing performance index when the system is adjusted as close to the model possible.

$I_{min\ min}$ is the value of the minimizing performance index for the model when it is ideal in the performance index sense.

This result is of questionable value because it establishes an upper bound value which is too large to be of practical use.

Rekasius [81] has suggested the following minimizing performance index.

$$I_k = \int_0^\infty \left[x^2 + \sum_{i=1}^{k} \tau_i^{\,2} \left(\frac{d^i x}{dt^i}\right)^2 + 2x \sum_{i=2}^{k} \tau_i \frac{d^i x}{dt^i} + \right.$$

$$\left. 2 \sum_{i=1}^{k} \sum_{j=i+2}^{k} \tau_i \tau_j \frac{d^i x}{dt^i} \frac{d^j x}{dt^j} \right] dt \quad k < n \tag{4.31}$$

where n is the order of the actual system and k is the order of the ideal model. x(t) is system error, which is defined as the difference between the desired value of the steady state response C_{ss} and the actual response of the closed loop system. If the system is asymptotically stable (and this is true for a stable linear system) the performance index becomes

$$I_k = \int_0^\infty \left[x + \sum_{i=1}^{k} \tau_i \frac{d^i x}{dt^i} \right]^2 dt + \tau_1 x^2(0) +$$

$$+ \sum_{i=1}^{k} \tau_i \, \tau_{i+1} \frac{d^i x(0)}{dt^i} \qquad k < n \tag{4.32}$$

Synthesis Procedure

The model is described by the characteristic equation

$$x + \tau_1 \frac{dx}{dt} + \dots + \tau_k \frac{d^k x}{dt^k} = 0 \tag{4.33}$$

The minimum value of this P. I. Equation (4.32) corresponds to the closed loop transfer function for the model.

$$\frac{C(s)}{R(s)} = \frac{1}{\tau_k s^k + \tau_{k-1} s^{k-1} + \dots + \tau_1 s + 1} \tag{4.34}$$

Hence this P. I. is applicable only if the model can be described by a unity numerator equation, which, in general, is not the case. To

specify the performance index in detail the model response must be described in the form

$$C(t) = C_{ss} - \sum_{i=1}^{k} A_i \, e^{-t/\tau_i} \qquad (4.35)$$

or

$$x(t) = \sum_{i=1}^{k} A_i \, e^{-t/\tau_i} \qquad (4.36)$$

From this equation the τ_i's are obtained. It may be necessary to resort to a graphical technique (Storer [82] , pp. 303-315) to obtain a mathematical description of the system.

Next to determine the performance index let

$$I_k = \lim_{t \to \infty} \int_0^t W(x, \frac{dx}{dt}, \ldots, \frac{d^k x}{dt^k}) \, dt = \lim_{t \to \infty} V(t) - V(o) \qquad (4.37)$$

In order to evaluate $V(t)$ one may assume it to be of the quadratic form

$$V = a_{11}x^2 + \sum_{j=2}^{n} a_{1j} \, x \, \frac{d^{j-1}x}{dt^{j-1}} + \sum_{i=2}^{n} \sum_{j \geq i}^{n} a_{ij} \, \frac{d^{i-1}x}{dt^{i-1}} \, \frac{d^{j-1}x}{dt^{j-1}} \qquad (4.38)$$

Since

$$W = \frac{dV}{dt} \qquad (4.39)$$

it is necessary to differentiate the V function and replace $\frac{d^n x}{dt^n}$ by the lower order derivative of x to obtain W. The characteristic equation of the system is

$$\frac{d^n x}{dt^n} + b_{n-1}\frac{d^{n-1}x}{dt^{n-1}} + \ldots + b_1 \frac{dx}{dt} + b_o x = 0 \qquad (4.40)$$

and the indicated process yields a W of the form

$$W = A_{11} x^2 + \sum_{j=2}^{n} A_{1j} x \frac{d^{j-1}x}{dt^{j-1}} + \sum_{i=2}^{n} \sum_{j}^{n} {}_{i} A_{ij} \frac{d^{i-1}x}{dt^{i-1}} \frac{d^{j-1}x}{dt^{j-1}} \qquad (4.41)$$

Equating A_{ij} to the corresponding terms of the I_k (performance index

equation) yields a set of

$$n + (n-1) + (n-2) + \ldots \ldots + 1$$

equations which are solved for all A_{ij}'s. V is now defined. For the

answer to have meaning it is necessary that the system be stable. Inci-

dentally the procedure used to evaluate I_k (as shown here) is identical

to the procedure of constructing Liapunov's functions for linear, autono-

mous systems 83 so that a check on stability is available. Routh's

criterion could also be used on the final system.

The procedure suggested for evaluating a performance index here is

the method of evaluating integrals by the use of exact differentials.

The method is illustrated by the following example.

Example 4.1

Consider a unity feedback system with the open loop transfer

function

$$G(s) = \frac{k}{s(1+s)^2} \qquad (4.42)$$

Let the model response be assumed as the following unity

numerator second order system. That is, it is assumed that the

step response of this second order model is given as ideal. In

general, the designer is free to pick the order of his model,

which in turn determines the order of the system, as the model must

be of order one less than the system.

$$\ddot{x} + 2\dot{x} + x = 0 \qquad (4.43)$$

By comparing equation (4.43) with equation (4.28), it is seen

that

$$\tau_1 = 2$$

$$\tau_2 = 1$$

From equation (4.31) the performance index becomes

$$I_2 = \int_0^{\infty} \left[x^2 + \tau_1^2 \dot{x}^2 + \tau_2^2 \ddot{x}^2 + 2\tau_2 \, x \ddot{x} \right] dt$$

$$= \int_0^{\infty} \left[x^2 + 4\dot{x}^2 + \ddot{x}^2 + 2x \ddot{x} \right] dt \qquad (4.44)$$

Since n=3 the V function equation (4.38) is

$$V = a_{11} x^2 + a_{12} x\dot{x} + a_{13} x\ddot{x} + a_{22} \dot{x}^2 + a_{23} \dot{x} \ddot{x} + a_{33} \ddot{x}^2 \qquad (4.45)$$

and

$$\frac{dV}{dt} = W = 2 a_{11} x \dot{x} + a_{12} x \ddot{x} + a_{12} \dot{x}^2 + a_{13} x \dddot{x}$$

$$+ a_{13} \dot{x} \ddot{x} + 2 a_{22} \dot{x} \ddot{x} + a_{23} \dot{x} \dddot{x} + a_{23} \ddot{x}^2 + 2a_{33} \ddot{x} \dddot{x} \qquad (4.46)$$

The closed loop transfer function for this system is

$$\frac{C(s)}{R(s)} = \frac{k}{s^3 + 2s^2 + s + k} \qquad (4.47)$$

and the characteristic equation is

$$\dddot{x} + 2\ddot{x} + \dot{x} + kx = 0 \qquad (4.48)$$

This equation is solved for $\dddot{x}$ and substituted into dV/dt to obtain

$$\frac{dV}{dt} = W = (-k \, a_{13})x^2 + (2a_{11} - a_{13} - k \, a_{23}) \, x \dot{x}$$

$$+ (a_{12} - 2a_{13} - 2k \, a_{33}) \, x \ddot{x} + (a_{12} - a_{23}) \, \dot{x}^2$$

$$+ (a_{13} + 2a_{22} - 2a_{23} - 2a_{33}) \, \dot{x} \ddot{x} + (a_{23} - 4a_{33}) \, \ddot{x}^2 \qquad (4.49)$$

Comparing this with the integrand of I_2 one may write

$$-k\, a_{13} = 1 \qquad\qquad a_{12} - 2a_{13} - 2k\, a_{33} = 2$$

$$a_{12} - a_{23} = 4 \qquad\qquad 2a_{11} - a_{13} - k\, a_{23} = 0 \qquad\qquad (4.50)$$

$$a_{23} - 4a_{33} = 1 \qquad\qquad a_{13} + 2a_{22} - 2a_{23} - 2a_{33} = 0$$

Simultaneous solution of these yields

$$a_{11} = \frac{k^3 + 4k^2 + 3k + 2}{2\,(k^2 - 2k)} \qquad\qquad a_{22} = \frac{k^2 + 6k}{k^2 - 2k}$$

$$a_{12} = \frac{5k^2 - 8k + 8}{k^2 - 2k} \qquad\qquad a_{23} = \frac{k^2 + 4k + 4}{k^2 - 2k} \qquad\qquad (4.51)$$

$$a_{13} = \frac{-1}{k} = \frac{2 - k}{k^2 - 2k} \qquad\qquad a_{23} = \frac{1.5\,k + 1}{k^2 - 2k}$$

The initial conditions for a step are

$$x(0) = 1$$

$$\dot{x}(0) = 0$$

$$\ddot{x}(0) = 0$$

Then

$$I_2 = V(0) = -\frac{k^3 + 4k^2 + 3k + 2 \; x^2\,(0)}{2(k^2 - 2k)}$$

$$= \frac{k^3 + 4k^2 + 3k + 2}{4k - 2k^2} \qquad\qquad (4.52)$$

The minimum value of I_2 yields the optimum system and it is

$$k = .43$$

at

$$I_{2_{min}} = 3.04$$

The response of the system is shown in Fig. 4.12.

The numerical value of the performance index for the model is

$$I = 2.00.$$

In this example only one system parameter was varied. The rise time is slower than that of the model and the system has 20% overshoot. Better results can be obtained by allowing more parameters to vary.

In conclusion, the philosophy of this approach is quite interesting. It is believed that this method is a fruitful area for further research. It indicates the possibility for the development of indices of performance. It is important that investigations be continued in this area.

4.2.19 Bellman's Performance Index

Abstract and Conclusions

Bellman's criterion is a general formulation rather than a quanity which can be used for system specification and evaluation.

Discussion

Bellman [52] has proposed a general performance index of the form

$$PI = G\left[\bar{c}_o(t) - \bar{c}(t)\right] + H\left[\bar{m}(t)\right] \tag{4.53}$$

$\bar{c}_o(t)$ is a vector representing the desired state of the control system, $\bar{c}(t)$ is the output in vector form, and $\bar{m}(t)$ is the control or input vector. $G\left[\bar{c}_o(t) - \bar{c}(t)\right]$ is a function which measures the cost of deviation from the ideal or desired state and $H\left[\bar{m}(t)\right]$ is a function which is a measure of the cost of control. Performance indicies such as ISE, IAE, etc., are special cases of this general formulation.

4.2.20 Generalized Error Function (GEF)

Abstract and Conclusions

This criterion is not satisfactory for system specifications and evaluation at the present time.

Discussion

Spooner and Rideout [84] have worked with a performance index called the "generalized error funtion (GEF)" which is defined by equation (4.54)

$$GEF = \lim_{t \to \infty} \frac{1}{2T} \int_{-T}^{T} \left[e(t,\tau) \right]^2 dt \qquad (4.54)$$

where

$$e(t,\tau) = r(t-\tau) - c(t)$$

The delayed input $r(t-\tau)$ is a stationary random signal.

4.2.21 Glover's Performance Index

Abstract and Conclusion

This criterion is used with Stochastic inputs. Sufficient results are not available to recommend the use at this criterion.

Discussion

Glover [85] has proposed a criterion called "the mean weighted square error" for filters. The criterion is expressed mathematically as

$$PI = \lim_{t \to \infty} \frac{1}{2T} \int_{-T}^{T} \frac{\left[f(t) - f_d(t) \right]^2}{\left[f_d(t) \right]^2 + \delta^2} \qquad (4.55)$$

where $f(t)$ is the obtained function of time and $f_d(t)$ is the desired function of time. δ^2 dictates the lowest absolute accuracy of interest. In this criterion, error is weighted in a per cent manner, since error is divided by the desired value of the function. It is Glover's belief that it is more reasonable to consider error on a percentage basis than on an absolute basis, as would be obtained by the mean-square error criterion. The quanity δ keeps the integrand finite and is chosen small enough to be essentially zero.

Murphy and Bold [86] considered a mean weighted square error previous to Glover's work, but they used a deterministic function of t for

weighting.

This method is interesting and mathematically tractable if adequate assumptions can be made concerning the statistical properties of the input. Sufficient results are not available at this time to predict the usefulness of the criterion.

4.2.22 Zaborszky and Diesel's Performance Index

Abstract and Conclusions

This criterion can not be used at the present time for system specification and evaluation as desired by the Air Force.

Discussion

Zaborszky and Diesel [87] have proposed a generalized error criterion which can be used with deterministic and random inputs. Many other criteria are special cases of their criterion, which is

$$PI = \int_0^\infty \overline{F[e(t),t,v_1,v_2,\ldots,v_r]p(t)dt} \tag{4.56}$$

where the penalty function F is a function of $e(t)$, time t, and parameters associated with the system $v_1, v_2 \ldots\ldots\ldots, v_r$. The quantity $p(t)$ is the probability density function of the times elapsing from activating the system to all times of utilization of its output. The bar indicates an averaging process over an ensemble of different types of inputs.

4.3 Analytical Determination of Performance Indices

4.3.1 Introduction

With the exception of ISE (integral of error squared) most work reported in the literature on performance indices has been done with computers; typically, a simulated system or family of systems has been studied by varying parameters until optimum values are obtained. The reason for

this approach is that most criteria are very cumbersome to handle analytically. Some mathematical methods that are appropriate for performance index problems will be presented and illustrated with examples in this section.

The mathematical approach to determining performance index values can be used to obtain the performance index of a specific system configuration or for the purpose of determining the optimum value of system parameters. There are two cases to consider in the optimization problem

a. *Semi-free system configurations i.e. some of the system parameters are fixed.

b. Free system configuration i.e. all system parameters are variable. An optimum free configuration is the best possible in the performance index sense.

For fixed configurations the only information desired is the value of the performance index and possibly its rate of change as different parameters are varied. For free and semi-free configurations it is desirable to know the number of minimum points, the performance index at these points and the gradient near the points.

The material presented here is primarily intended for the class of integral performance indices of the form

$$\int_0^\infty t^n e^2(t)\, dt \qquad\qquad n = 0,1,2 - - - \qquad\qquad (4.57)$$

4.3.2 Parseval's Theorem

---

* Note that semi-free and free configurations are defined here. The definition differs from that of Newton, Gould and Kaiser [61] .

Parseval's theorem [61] is not directly applicable to the whole family of performance indices described by equation (4.57), but with modification it can be applied to the whole family. The main utility of Parseval's theorem is that it yields time domain answers from frequency domain information, making it unnecessary to obtain system error as a function of time. If system error is

$$e(s) = \frac{c(s)}{d(s)}$$

the performance index for the case where n = 0 is

$$PI = \int_0^\infty e(t)\, dt = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{c(s)\, c(-s)}{d(s)\, d(-s)}\, ds \qquad (4.58)$$

by Parseval's theorem.

The integral on the right side of equation (4.58) can be solved by determining the residues of the integrand and its solution has been tabulated [61] for equations up to the tenth order of the form

$$\frac{p(s)}{q(s)}$$

where p(s) and q(s) are polynomials. The solution is in terms of the coefficients of the polynomials. This makes it unnecessary to determine the roots of the polynomials.

Thus, it is easy to obtain the performance index value of a fixed configuration system where n = 0 in equation (4.57). However, when a system is to be optimized the computational difficulties may be formidable, since the solutions available are only a starting point for the optimization process. To determine analytically the optimum system parameters it is necessary to take the partial derivative of the performance index with respect to each parameter and set the resulting

equations equal to zero or use some other process. A naive approach is to vary one parameter at a time until the desired minimum is obtained. This may actually require less work than the analytical solution. If the optimum values are not finite or if they are zero, it is necessary to introduce constraints to obtain a non-trivial solution.

The optimization problem is easier for practical systems of a semi-free configuration than for free configurations because fewer parameters vary. Most practical systems will have semifree configurations because a motor or control surface etc. will have fixed characteristics. In many cases it is only necessary to vary the parameters of an equalizer.

A simple second order system will serve the purpose of illustrating the use of Parseval's theorem. Assume a system with the open loop transfer function.

$$A = \frac{K}{s(Ts + 1)} \qquad (4.59)$$

is to be optimized [61] . First, an equation of the error must be obtained, which, if $r(t)$ is a unit step, becomes

$$e(s) = \frac{ts+1}{ts^2 + s+1} \qquad (4.60)$$

Then from equation (4.58)

$$PI = \int_0^\infty e^2(t)\, dt = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left(\frac{T s + 1}{T s^2 + s + k}\right)\left(\frac{-T s + 1}{T s^2 - s + k}\right)ds \qquad (4.61)$$

$$= \frac{c_1^2 d_0 + c_0^2 d_2}{2\, d_0 d_1 d_2} \qquad (4.62)$$

where

$$d_2 = T \qquad\qquad c_1 = T$$

$$d_1 = 1 \qquad\qquad c_0 = 1$$

$$d_0 = K$$

Since, the general solutions are available for equations such as (4.61)

up to the tenth order, obtaining the solution is routine. The perform-

ance index becomes

$$PI = \frac{T + 1/K}{2} \qquad (4.63)$$

At this point in the analysis it is necessary to consider stability.

Since this method does not insure stability and Parseval's theorem is

applicable only for stable systems, the results are meaningless unless

they lead to a stable system. In this example it is obvious that K must

be positive to insure stability assuming T is a positive quantity.

From equation (4.63) it can be seen that the PI approaches infinity

as K approaches zero and approaches the value T/2 as K approaches infinity.

This difficulty can be overcome in this example by relating T and K to

$\mathcal{f}$ in the usual second order system terminology.

$$\mathcal{f} = \frac{1}{2(TK)^{1/2}} \qquad (4.64)$$

$$\omega_n = (\frac{K}{T})^{\frac{1}{2}} \qquad (4.65)$$

Then, from equation (4.63)

$$PI = \frac{1}{2} (\frac{1}{4 \mathcal{f}^2 K} + \frac{1}{K})$$

$$= \frac{1}{2\omega_n} (\frac{1}{2\mathcal{f}} + 2\mathcal{f})$$

$$\frac{d(ISE)}{d\mathcal{f}} = \frac{1}{2\omega_n}(- \frac{1}{2\mathcal{f}^2} + 2) = 0 \qquad (4.66)$$

Thus, the optimum second order system in an ISE sense has a $\mathcal{f} = 0.5$.

In this system it was only necessary to relate T and K to $\mathcal{f}$ to ob-

tain closed contours of performance index versus system parameters. In more complicated systems the procedure may not be so obvious and this becomes a limiting factor in analytical design. For example, it is obvious that the ideal system in the sense of equation (4.57) should have infinite bandwidth and hence, the analytical solution may force time constants to zero if proper constraints are not included to obtain nontrival solutions. In practical semifree configurations the problem is less difficult than for free configuration systems because only a few zeros and poles are varied. Even in this case, however, it may be necessary to add constraints.

4.3.3 Analytical Solution of ITSE

Westcott [88] has shown that Parseval's theorem can be extended to solve for the integral of Time Multiplied by Squared Error. This makes it possible to utilize the solutions that have been tabulated for the use of Parseval's theorem. The following derivation follows Westcott. It is desired to express

$$ITSE = \int_0^\infty t\, e^2(t)\, dt \tag{4.67}$$

in terms of the coefficients of the polynomials $p(s)$ and $q(s)$ of the error transfer function

$$e(s) = \frac{p(s)}{q(s)} \tag{4.68}$$

Let the Laplace transform of $e^2(t)$ be $F(\sigma_1)$, i.e.

$$F(\sigma_1) = \int_0^\infty e^2(t)\, e^{-\sigma_1 t}\, dt \tag{4.69}$$

Then

$$\int_0^\infty t \, e^2(t) \, dt = - \lim_{\sigma_1 \to 0} \frac{\partial}{\partial \sigma_1} \int_0^\infty e^2(t) \, e^{-\sigma_1 t} \, dt \qquad (4.70)$$

$$= - \lim_{\sigma_1 \to 0} \frac{\partial}{\partial \sigma_1} \left[F(\sigma_1) \right] \qquad (4.71)$$

By the Laplace transform theory it may be shown that

$$F(\sigma_1) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} e(s) \, e(\sigma_1 - s) \, ds \qquad (4.72)$$

and it follows that

$$ITSE = - \lim_{\sigma_1 \to 0} \frac{\partial}{\partial \sigma_1} \left[\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} e(s) \, e(\sigma_1 - s) \, ds \right] \qquad (4.73)$$

This integral is not symmetrical, as desired, but may be made symmetrical

by choosing the proper path of integration. After choosing the proper

path of integration and making a change of variable the desired

symmetrical form is

$$ITSE = - \lim_{\sigma \to 0} \frac{\partial}{\partial \sigma} \left[\frac{1}{4\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} e(s+\sigma) e(\sigma - s) \, ds \right] \qquad (4.74)$$

To illustrate the method two examples will be presented. ([88], p. 479).

Example 4.2

Consider a system with the error transfer function

$$e(s) = \frac{d_o}{a_o s + a_1} \qquad (4.75)$$

Then

$$ITSE = - \lim_{\sigma \to 0} \frac{\partial}{\partial \sigma} \left[\frac{1}{4\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{d_o^2 \, ds}{\left[a_o(s+\sigma) + a_1 \right] \left[a_o(\sigma - s) + a_1 \right]} \right] \qquad (4.76)$$

By determining the residue at the pole

$$s = - (\frac{a_1}{a_0} + \sigma)$$ (4.77)

the quantity

$$ITSE = - \lim_{\sigma \to 0} \frac{\partial}{\partial \sigma} \left[\frac{d_0^2}{4a_0(a_0\sigma + a_1)} \right]$$ (4.78)

is obtained. Then

$$ITSE = - \lim_{\sigma \to 0} \frac{d_0^2}{4(a_0\sigma + a_1)^2}$$ (4.79)

after taking the indicated partial derivative. Upon taking the limit the answer becomes

$$ITSE = - \frac{d_0^2}{4a_1^2}$$ (4.80)

Example 4.3

Consider the system defined by equation (4.59) for the second example. Using the procedure indicated above it is found that

$$ITSE = \frac{1}{4K^2} + \frac{T^2}{2}$$ (4.81)

After normalizing to make the system a unity numerator type (i.e. T=K) and differentiating ITSE to obtain the optimum, it is found that

$$\mathcal{T} = 0.658$$

Westcott's work may be extended to InTSE. Using the method outlined by Westcott

$$\int_0^\infty t^n e^2(t)\, dt = (-1)^n \lim_{\sigma_1 \to 0} \frac{\partial^n}{\partial \sigma_1^n} \left[\int_0^\infty e^2(t)\, e^{-\sigma_1 t}\, dt \right]$$ (4.82)

$$= (-1)^n \lim_{\sigma_1 \to 0} \frac{\partial^n}{\partial \sigma_1^n} \left[F(\sigma_1) \right]$$ (4.83)

where $F(\sigma_1)$ is the Laplace transform of $e^2(t)$.

It follows from the previous derivation that

$$\text{InTSE} = \frac{(-1)^n}{2} \lim_{\sigma \to 0} \frac{\partial^n}{\partial \sigma^n} \left[\frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} e(s+\sigma)e(\sigma - s)ds \right] \quad (4.84)$$

4.3.4 Analytical Solution for ISTSE

The results of the last section can be used to obtain ISTSE. When

n equals 2, equation (4.82) becomes

$$\text{ISTSE} = \frac{1}{2} \lim_{\sigma \to 0} \frac{\partial^2}{\partial \sigma^2} \left[\frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} e(s+\sigma)e(\sigma - s)ds \right] \quad (4.85)$$

Example 4.4

Consider a system which has the error transfer function

$$e(s) = \frac{d_o}{(a_o s + a_1)} \quad (4.86)$$

Then

$$\text{ISTSE} = \lim_{\sigma \to 0} \frac{\partial^2}{\partial \sigma^2} \left[\frac{1}{4\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{d_o^2 \, ds}{\left[a_o(s+\sigma)+a_1\right]\left[a_o(\sigma - s)+a_1\right]} \right] \quad (4.87)$$

$$= \lim_{\sigma \to 0} \frac{\partial^2}{\partial \sigma^2} \left[\frac{d_o^2}{4 a_o (a_o \sigma + a_1)} \right]$$

$$= \lim_{\sigma \to 0} \frac{a_o d_o^2}{2 (a_o \sigma + a_1)^3}$$

$$= \frac{a_o d_o^2}{2a_1^3} \quad (4.88)$$

As an alternate method to obtain ISTSE it is noted that

$$\mathcal{L}\left\{t\left[e(t)\right]\right\} = -\frac{d\,[e(s)]}{ds} \quad (4.89)$$

from the Laplace transform theory. Thus it is only necessary to differ-

- 104 -

entiate $-e(s)$ and insert it in the integral of Parseval's theorem instead
of $e(s)$ to make use of tabulated results. That is,

$$\text{ISTSE} = \int_0^\infty t^2 e^2(t)\, dt = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} \frac{de(s)}{ds}\frac{de(-s)}{ds}\, ds \tag{4.90}$$

The system defined by equation (4.59) will be used to illustrate this
method.

$$-\frac{de}{ds} = \frac{T^2 s^2 + 2\,Ts + (1-TK}{T^2 s^4 + 2Ts^3 + (1+2T^2)\,s^2 + 2sT+T^2} \tag{4.91}$$

Using solution tables for ISE and setting T equal to K as done previously
yields

$$\text{ISTSE} = \frac{4T^6 + T^4 - T^2 + 1}{4T^3} \tag{4.92}$$

Then

$$\frac{d(\text{ISTSE})}{dT} = \frac{12T^6 + T^4 + T^2 - 3}{4T^4} = 0 \tag{4.93}$$

and

$$T = .75$$

which yields

$$\mathcal{J} = .666$$

4.3.5 Analytical Solution of Performance Indices Using Liapunov V Function

Another analytical method for determining the value of a performance
index is available from the relationship of the V and W functions used in
the second method of Liapunov (see Final Report Vol. 3). By definition

$$W(x_1, x_2, \ldots, x_n) = \frac{dV}{dt} \tag{4.94}$$

Integrating this expression with respect to time yields

$$V(x_1, x_2, \ldots, x_n) - V(x_{10}, x_{20}, \ldots, x_{no}) = \int_0^t W(x_1, x_2, \ldots, x_n)\, dt \tag{4.95}$$

where the zero subscripts indicate the initial values of the state varia-
ble x_i at t=0, i.e. the initial conditions. If time is allowed to approach
infinity in equation (4.95) (the case for most integral performance indices),
the equation becomes

$$V(x_{10}, x_{20}, \ldots, x_{no}) - \lim_{t \to \infty} V(x_1, x_2, \ldots, x_n) = - \int_0^\infty W(x_1, x_2, \ldots x_n) dt \tag{4.96}$$

By definition, in an asymptotically stable system

$$\lim_{t \to \infty} x_i = 0 \qquad\qquad i = 1, 2, \ldots, n \tag{4.97}$$

and

$$V(o, \ldots, o) = 0 \tag{4.98}$$

Then

$$V(x_{10}, x_{20}, \ldots, x_{no}) - \int_0^\infty W(x_1, x_2, \ldots, x_n) dt \tag{4.99}$$

If the state variables x_i represent system error and its n-1 time
derivatives, and if W is a positive definite quadratic form, equation (4.99)
is an integral error type performance index, where

$$PI = V(x_{10}, x_{20}, \ldots, x_{no}) \tag{4.100}$$

In general, any Liapunov function which has a negative definite or
negative semidefinite derivative can be used as a performance index.

For time weighted performance indices (eq. ITE) it is necessary to
assume the proper V function e.g.

$$V = V_1(e, \dot{e}, \ldots) + tV_2(e, \dot{e}, \ldots) \tag{4.101}$$

No examples of this type are available at the present time.

Example 4.5

Consider the following example of this method where a unity

feedback system has the open loop transfer function

$$G(s) = \frac{K}{s(s+a)} \qquad (4.102)$$

and where it is desired to obtain

$$P.I. = \int_0^\infty e^2(t)dt \qquad (4.103)$$

The error transfer function is

$$\frac{e(s)}{R(s)} = \frac{1}{1+ \dfrac{K}{s(s+a)}} \qquad (4.104)$$

If a step function R (the input) is removed at t=0 the following is obtained from equation (4.104)

$$\ddot{e}(t) + a\,\dot{e}(t) + Ke(t) = 0 \qquad t>0 \qquad (4.105)$$

Let

$$x_1 = e(t) \qquad (4.106)$$

$$x_2 = \dot{e}(t) \qquad (4.107)$$

then

$$\dot{x}_1 = x_2 \qquad (4.108)$$

$$\dot{x}_2 = -a\,x_2 - Kx_1 \qquad (4.109)$$

Assume the general quadratic form

$$V = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 \qquad (4.110)$$

From equation (4.94) and (4.10)

$$W = 2a_{11}x_1\dot{x}_1 + 2a_{12}(\dot{x}_1x_2 + x_1\dot{x}_2) + 2a_{22}x_2\dot{x}_2 \qquad (4.111)$$

and

$$W = (-2Ka_{12})x_1^2 + (2a_{11} - 2aa_{12} - 2Ka_{22})x_1x_2 + (2a_{12} - 2aa_{22})x_2^2 \qquad (4.112)$$

Since W must be the integrand of equation (4.103) it is neccessary that

$$W = x_1^2 = e^2(t) \tag{4.113}$$

Using equation (4.112) and the above condition yields

$$a_{12} = -\tfrac{1}{2}K$$

$$a_{22} = -\tfrac{1}{2}Ka \tag{4.114}$$

$$a_{11} = -\frac{k+a^2}{2Ka}$$

Then

$$V = -\frac{K+a^2}{2Ka} x_1^2 - \frac{x_1 x_2}{K} - \frac{x_2^2}{2Ka} \tag{4.115}$$

To find the initial values of x_1 and x_2 it is only necessary to apply the initial value theorem to equation (4.104) and the derivative of equation (4.104) and obtain

$$x_1(0) = 1 \tag{4.116}$$

$$x_2(0) = 0 \tag{4.117}$$

From equation (4.99) and (4.115) it is found that

$$PI = \frac{K + a^2}{2Ka} \tag{4.118}$$

For this simple problem it would be easier to use Parseval's theorem, however, for more general performance indices (e.g. $PI = f(e, \dot{e} \text{---})$) this method is superior.

BIBLIOGRAPHY

1. J. E. Gibson, E. S. McVey, C. D. Leedham, Z. V. Rekasius and R. Sridhar, "Specification and Data Presentation in Linear Control Systems," vol. 1 Final Report on project AF 29(600)-1933, Purdue University, 1960.

2. J. T. Tou, Digital and Sampled Data Control Systems, McGraw-Hill Book Co., New York, N. Y., 1959.

3. J. G. Truxal, Automatic Feedback Control System Synthesis, McGraw-Hill Book Co., New York, N. Y., 1955.

4. J. R. Ragazzini and G. F. Franklin, Sampled Data Control Systems, McGraw-Hill Book Co., New York, New York, 1958.

5. W. K. Linvill, "Sampled-data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation," Trans. AIEE, vol. 70, part 2, p. 179, 1951.

6. H. S. Black, Modulation Theory, D. Van Nostrand and Company Inc., New York, New York, 1960.

7. W. L. Nelson, "Pulse Width Control of Sampled Data Systems" Technical Report No. T-35/B Columbia University, New York, New York.

8. R. B. Marshall, Measurements in Electrical Engineering, vol 1, John S. Swift Co., Cincinnati, Ohio, 1948.

9. J. R. Ragazzini and L. A. Zadeh, "The Analysis of Sampled Data Systems," Trans. AIEE, vol. 71 part 2, p. 225, 1952.

10. E. I. Jury, "Analysis and Snythesis of Sampled Data Control Systems," Trans. AIEE, vol 73, part 1, p. 332, 1954.

11. R. H. Barker, "The Pulse Transfer Function and its Application to Sampling Servo Systems," Proc. I.E.E. (England) vol. 99, p. 1, 1952.

12. E. I. Jury, "Synthesis and Critical Study of Sampled Data Control Systems," Trans. AIEE, vol. 75, part 2, p. 141, 1956.

13. E. O. Gilbert, "A Method for the Symbolic Representation and Analysis of Linear Periodic Feedback Systems," Trans. AIEE, vol. 78, part 2, p. 512, Jan., 1960.

14. F. Kurzweil, Jr., "The Analysis and Synthesis of Nonlinear Continuous and Sampled-Data Systems," Technical Report No. 2101-1, November, 1959, Stanford Electronics Laboratories, Stanford University, California.

15. R. E. Kalman and J. E. Bertram, "A Unified Approach to the Theory of Sampling Systems," Journal of the Franklin Institute, vol. 267, p. 405, 1959.

16. E. I. Jury, "Hidden Oscillations in Sampled Data Control Systems," Trans. AIEE, vol. 75, part 2, p. 391, 1956.

17. G. W. Johnson, Lindorff and Nordling, "Extension of Continuous Data Design Technique to Sampled Data Control Systems," Trans. AIEE, vol. 74, part 2, p. 252, 1955.

18. Schmidt, "Application of Continuous System Design Concepts to the Design of Sampled Data Systems," Trans. AIEE, vol. 78, part 2, p. 74, 1959.

19. R. G. Brown and G. J. Murphy, "An approximate Transfer Function for the Analysis and Design of Pulsed Servos," Trans. AIEE, vol. 71, part 2, p. 436, 1952.

20. H. Nyquist, "Certain topics in Telegraph Transmission Theory" AIEE Transaction p. 617, April 1928.

21. C. E. Shannon, "Communication in the presence of Noise," Proc. IRE, vol. 37, No. 1, Jan. 1949.

- 110 -

22. E. Hille, <u>Analytic Function Theory</u> vol. 1, Ginn & Co., Boston, Mass.

23. H. Chestnut and R. W. Mayer, <u>Servomechanism and Regulating System Design, vol. II,</u> John Wiley and Sons, Inc., New York, N. Y., 1955.

24. E. I. Jury, "The Correlation Between Root Locus and Transient Response of Sampled Data Control Systems," Trans. AIEE, vol. 75, part 2, Jan., 1956.

25. M. Mori, "Root Locus Method of Pulse Transfer Function for Sampled Data Control Systems," Trans IRE, PGAC 3, November 1957.

26. E. I. Jury, "The Effect of Pole and Zero locations on the Transient Response of Sampled Data Systems," Trans. AIEE, vol 74, part 2, 1955.

27. J. J. Stoker, <u>Nonlinear Vibrations,</u> Interscience publisher. New York, 1950

28. E. L. Ince, <u>Ordinary Differential Equations,</u> Dover, 1956.

29. R. E. Bellman, Stability Theory of Differential Equations. McGraw-Hill Book Co., 1953.

30. L. A. Zadeh, "On Stability of Linear Varying Parameter Systems," J.A.P. vol. 22, pp. 402 - 405, April 1951.

31. H. A. Antosiewicz, "A survey of Liapunov's second method," Contributions to the theory of nonlinear Oscillatrions IV, Annals of Mathematics Studies, No. 41, Princeton University Press, 1958.

32. R. E. Kalman, and J. E. Bertram, "Control System Analysis and Design via the "Second Method" of Liapunov," Journal of Basic Engineering, Trans. ASME, part D, pp. 371-393, June 1960.

33. J. L. Massera, "Contributions to stability theory," Annals of Mathematics, vol. 64, No. 1, pp. 182 - 206, July 1956.

34. F. R. Gerardi, "Application of Mellin and Hankel Transforms to Networks with Time Varying Parameters," PGCT, vol. CT6, No. 3, p. 197, June 1959.

35. M. J. Kirby, "Stability of Servomechanisms with Linearly Varying Elements," Trans AIEE, vol 69, pt. 2, 1950, pp. 1662 - 1668.

36. M. J. Kirby and R. M. Giulianelli, "Stability of Varying Element Servomechanisms with Polynomial Coefficients," Trans. AIEE vol. 70, part 2, pp. 1447 - 1451, 1951.

37. J. Brodin, "Analysis of Time Dependent Linear Networks," PGCT March 1955, pp. 12 - 16.

38. W. R. Bennett, "Steady State Transmission Through Networks Containing Periodically Operated Switches," PGCT March 1955, pp. 17 - 21.

39. V. V. Karamyshkin, "Transformation of a Linear Differential Equation with Polynomial Coefficients into an Integral Equation with the aid of Operational Calculus," "Prekladnays Mathematica, i Mexanike" vol 22, No. 4, pp. 774 - 776, 1958

40. C. A. Desoer, "Steady State Transmission Through a Network Containing a Single Time Varying Element," PGCT 1959, pp. 244 - 252.

41. L. A. Pipes, "Matrix Analysis of Linear Time Varying Systems," PGCT December 1953, pp. 91 - 104.

42. L. A. Zadeh, "Frequency Analysis of Variable Networks," Proc. of IRE, vol. 38, pp. 291 - 299, March 1950.

43. L. A. Zadeh, "The Determination of Impulsive Response of Variable Networks," J.A.P., vol. 21, pp. 642 - 645, July 1950.

44. L. A. Zadeh, "Initial Conditions in Linear Varying Parameter Systems," J.A. P., vol 22, pp. 782 - 786, July 1951.

45. L. A. Zadeh, "Circuit Analysis of Linear Varying Parameter Networks," J.A.P., vol. 21, pp. 1171 - 1177, Nov. 1950

46. J. A. Aseltine, "A Transform Method for Linear Time Varying Systems," J.A.P., vol. 25, No. 6, pp. 761, June 1954.

47. I. G. Malkin, "Theory of Stability of Motion," AEC Translation 3352 USDC, 1958.

48. G. P. Szego, "A Stability Investigation of Time Varying Linear Systems," To be presented at the AIEE general meeting, Summer 1961.

49. J. Matyash, "The Analog Computer Setup Procedure for Solving Linear Differential Equations with Variable Coefficients," Automatika i Telemexanika, vol. 10, No. 7.

50. J. H. Laning and R. H. Battin, "Random Processes in Automatic Control," Chapter 6. McGraw-Hill Book Co., New York, New York, 1956.

51. G. W. Anderson, J. A. Aseltine, A. R. Mancini, and C. W.Sarture, "A Self-Adjusting System for Optimum Dynamic Performance," IRE Convention Record, part 4, AC 1958, pp. 182 - 190, 1958.

52. R. E. Bellman, Dynamic Programming, Princeton University Press, Princeton, N. J., 1957.

53. D. Graham and R. C. Lathrop, "The Synthesis of 'Optimum' Transient Response: Criteria and Standard Forms," Transactions AIEE, vol. 72, part 2, A & I, pp. 273 - 288, 1954.

54. M. A. Aizerman, Lectures on the Theory of Automatic Control, in Russian, second ed., Gostekizdat, pp. 302 - 320, 1958.

55. R. L. Cosgriff, Nonlinear Control Systems, John Wiley and Sons, Inc., New York, N. Y., 1958.

56. H. H. Skilling, Transient Electric Circuits, McGraw-Hill Book Co., New York, N. Y., 1952.

57. R. C. Oldenbourg and H. Sartorius, The Dynamics of Automatic Controls, The American Society of M. E., New York, New York, 1948.

58. P. T. Nims, "Some Design Criteria for Automatic Controls," Transactions AIEE, vol. 70, part 1, pp. 606 - 611, 1951.

59. F. C. Fiechesen and T. M. Stout, "Analogue Methods for Optimum Servomechanism Design," Transactions AIEE, A & I, vol. 71, pp. 244-250, 1952.

60. A. C. Hall, The Analysis and Synthesis of Linear Servomechanisms, Technology Press, MIT, 1947.

61. G. C. Newton, L. A. Gould, and J. R. Kaiser, Analytical Design of Linear Feedback Controls, John Wiley and Sons, Inc., New York, N. Y. 1957.

62. R. N. Clark, "Integral of the Error Squared as a Performance Index for Automatic Control Systems," Trans. AIEE, A and I, pp. 467 - 471, Jan. 1961.

63. H. M. James, N. V. Nichols, and R. S. Phillips, Theory of Servomechanisms, McGraw-Hill Book Co., New York, N. Y., 1947.

64. J. G. Truxal (Editor), Control Engineers Handbook, McGraw-Hill Book Co., New York, New York, 1958.

65. N. Weiner, "The Extrapolation, Interpolation, and Smoothing of Stationary Time Series," John Wiley and Sons, Inc., New York, 1948.

66. E. A. Guillemin, Communication Network, vol. 1, John Wiley and Sons, Inc., New York, N. Y., 1931.

67. J. W. Froggatt, "Investigation of Several Criteria for the Synthesis of Optimum Transient Response of Servomechanism Systems of Higher Orders," Air University, U.A.A.F. Inst. of Tech., Wright-Patterson AF Base, Ohio, 1954.

68. H. S. Black, "Stabilized Feedback Amplifiers," Transactions AIEE, vol. 53, pp. 114 - 120, 1934.

69. H. Nyquist, "Regeneration Theory," B.S.T.J., vol. 11, pp. 126 - 147, 1932.

70. G. E. Happell and W. M. Hesselberth, Engineering Electronics, McGraw-Hill Book Co., New York, N. Y., 1953.

71. D. Graham and R. C. Lathrop, "The Transient Performance of Servo-mechanisms with Derivative and Integral Control," Transactions AIEE, vol. 72, part 2, A & I, pp. 10 - 17, 1954.

72. D. Graham and R. C. Lathrop, "The Influence of Time Scales and Gain on Criteria for Servomechanism Performance," Transactions AIEE, vol. 72, part 2, A & I, pp. 153 - 158, 1954.

73. Gibson, J., McVey, E., "Multidimensional Adaptive Control," Proc. NEC, vol. 15, Oct. 1959.

74. J. E. Gibson, "Making Sense out of the Adaptive Principle," Control Engineering, vol 7. pp. 113 - 117, August, 1960.

75. R. C. Stone, "A Mathematical Method for Optimizing the Gains of A Linear System," Aero Report 49174-21 TR1, Minneapolis Honeywell Reg. Co., Minneapolis, Minn.; Feb., 1957.

76. J. Crow, "An Integral Criterion for Optimizing Duplicator Systems on the Basis of Transient Response," Doctoral Dissertation, Washington University, St. Louis, Mo., 1956.

77. V. C. Rideout and W. C. Schultz, "Control System Performance Measured: Past, Present, and Future," IRE, PGAC, vol. AC-6m pp. 22 - 35: Feb. 1961.

78. H. M. Paynter, "On An Analogy Between Stochastic Process and Monatone Dynamic Systems," Regelvngstechnik Modern Theoriem und Ihre Vernendbor-heir, Oldenbavng, Munich, Germany; 1957.

79. T. P. Goodman and R. H. Hillsley, "Continuous measurement of Character-istics of Systems with Random Inputs: A step toward Self-optimizing Control," Trans. ASME, vol. 80, pp. 1839 - 1848; Nov. 1958.

80. R. F. Kalman and R. W. Koepeke, "Optimal Synthesis of Linear Sampling Systems Using Generalized Performance," Trans. ASME, vol. 80, pp. 1820 - 1826; Nov. 1958.

81. Z. V. Rekasius, "A General Performance Index for Analytical Design of Control Systems," to be presented at JACC, June 1961.

82. J. E. Storer, Passive Network Synthesis, McGraw-Hill Book Co., New York, New York, pp. 303 - 315, 1959.

83. J. E. Gibson, et. al., "Stability of Nonlinear Control Systems by The Second Method of Liapunov," vol 111 of Final Report on project AF 29(600)-1933, Purdue University, May 1961.

84. M. G. Spooner and V. C. Rideout, "Correlation Studies of Linear and Nonlinear Systems," paper presented at The National Electronics Conference Chicago, Illinois, October, 1956.

85. C. C. Glover, "Optimon Prediction with a Mean Weighted Square Error Criterion," IRE, PGAC, pp. 43 - 48, Feb., 1961.

86. G. J. Murphy and N. T. Bold, "Optimization Based on a Square Error Criterion With an Arbitrary Weighting Function," IRE PGAC, vol. AC-5, pp. 24 30; Jan., 1960.

87. J. Zaborszky and J. W. Diesel, "Probabilistic Error as a Measure of Control System Performance," Trans. AIEE, A and I, no. 43, pp. 163 - 168, July 1959.

88. Westcott, "The Minimum-Moment-of-Error-Squared Criterion," Proc. IEE, part 2, p. 480; 1954.

APPENDIX A

A DISCUSSION OF SAMPLING FREQUENCY

A.1 Introduction

This appendix is included to supplement the material used in making the recommendations of Chapter 2 (Sampled Data Systems).

A question often asked by engineers when dealing with sampled data systems is the following: "Given a system, is there a sampling frequency above which the system can be considered continuous for most purposes?" This appendix provides a partial answer to this specific question, which was put to the project staff by Mr. J. H. Gengelbach, the initiator of this study.

The research project, of which this volume is a part, is not designed as a project in basic sampled-data system research. The examples choosen in the following discussion are, therefore, restricted to the simplest examples (second order systems) possible.

A.2 Discussion

One of the principle characteristics associated with a sampled data system is the periodicity with which samples are obtained from continuous data. It is clear that the sampling frequency will affect the performance of a given system, and therefore questions will arise naturally as to what is the best sampling frequency. Alternatively, given a sampling frequency and a system, what performance can be expected, and how shall it be assessed. It is recognized that the frequency at which samples can be obtained is often outside a designer's control and dictated by external factors. The purpose of this appendix

will be to investigate the performance of a system as a function of

sampling period, T, (sampling frequency, $\omega_s = \frac{2\pi}{T}$), assuming this

quantity can be varied continuously from zero to a high figure, and

to answer the question posed above.

A standard must be chosen to which the performance of the sampled

data system can be compared. It is suggested, intuitively, that the

ultimate performance of a given sampled system is the performance of

the same system, but without the sampler and associated circuits

present. For example, consider the error sampled system shown in

Figure A.1. There the sampler is considered to be represented by an

impulse modulation device, and the zero order hold circuit, G_{HO}, is

introduced to make the system realistic. After removing the sampler

and hold circuit, the system becomes a continuous system as in Figure

A.2. This system can, in turn, be reduced mathematically to a transfer

function which has the familiar form:

$$\frac{C(s)}{R(s)} = \frac{K/T_a}{s^2 + \frac{1}{T_a}s + \frac{K}{T_a}} = \frac{\omega_o^2}{s^2 + 2\mathcal{f}\omega_o s + \omega_o^2} \tag{A.1}$$

where $KT_a = \frac{1}{4\mathcal{f}^2}$ and $\frac{K}{T_a} = \omega_o^2$

This expression is that of a second order system and the response

at the output terminals to a step of magnitude A at the input is:

$$c(t) = A - 2\sqrt{KT_a}\, A\, \epsilon^{\frac{-t}{2T_a}}\, \sin\left[\frac{\sqrt{4KT_a-1}\,t}{2T_a} + \phi\right] \tag{A.2}$$

where $\phi = \tan^{-1}\sqrt{4KT_a - 1}$

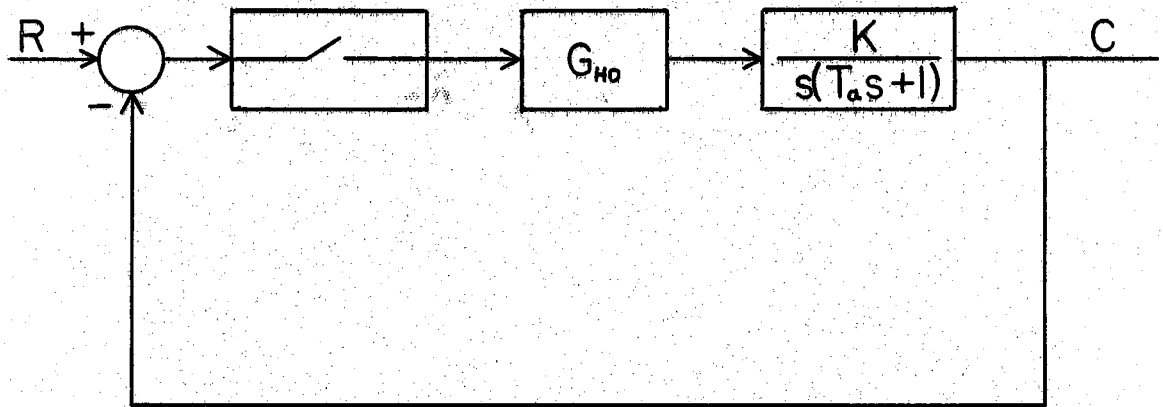provided $0 < \mathcal{f} < 1$ or $0 < \frac{1}{4KT_a} < 1$

Figure A-1

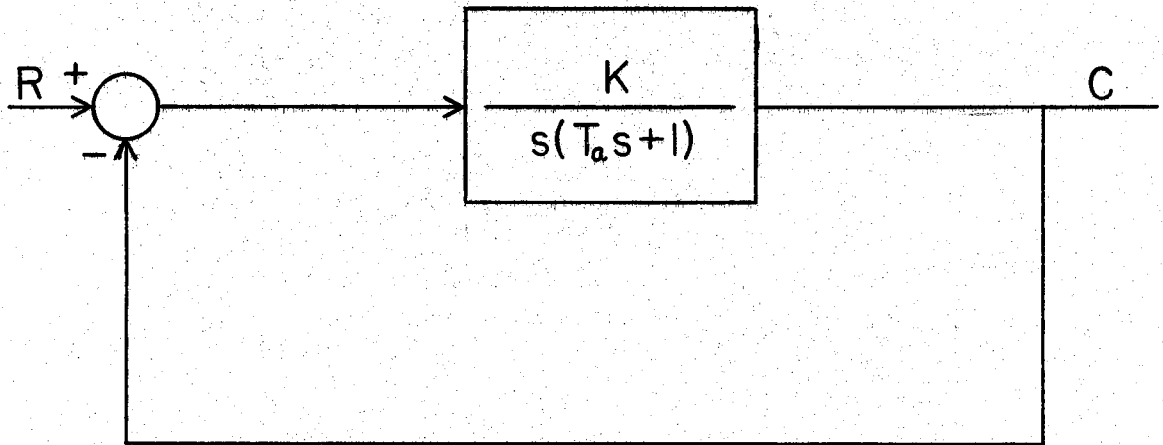An Error Sampled Second Order System

Figure A-2

The Equivalent Continuous System

The step response of the system shown in Figure A.1, which includes the sampling device, is to be compared with equation (A.2). The two responses could be of the form shown in Figure A.3.

The work contained in this volume on performance indicies suggests that an index could be used for comparison purposes and that the index should be representative of the area enclosed between the two curves in Figure A.3. Both positive and negative area will be generated, but the absolute area is the meaningful quantity, and a squared index will be choosen. The simplest suitable index is:

$$P.I. = \int_0^\infty \left[c_c(t) - c_s(t) \right]^2 dt \qquad (A.3)$$

where $c_c(t)$ ... the output from the continuous system to a step.

$c_s(t)$ ... the output from the sampled system to the same step.

For convenience, let

$$c_c(t) - c_s(t) = c_e(t).$$

Then

$$P.I. = \int_0^\infty c_e^2(t) \, dt. \qquad (A.4)$$

Recalling Parseval's Theorem $\left(ref. [25], p. 43 \right)$ the expression may also be put in the form:

$$P.I. = \int_0^\infty c_e^2(t) \, dt = \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} C_e(s) C_e(-s) \, ds \qquad (A.5)$$

where s is the complex variable associated with the Laplace Transform.

It is recognized that other performance indicies exist that are suitable for the comparison of these two systems, for example:

$$P.I. = \sum_{n=0}^\infty \left[c_c(nT) - c_s(nT) \right]^2 \qquad (A.6)$$

but attention will be focussed here on that index in equation (A.5) which
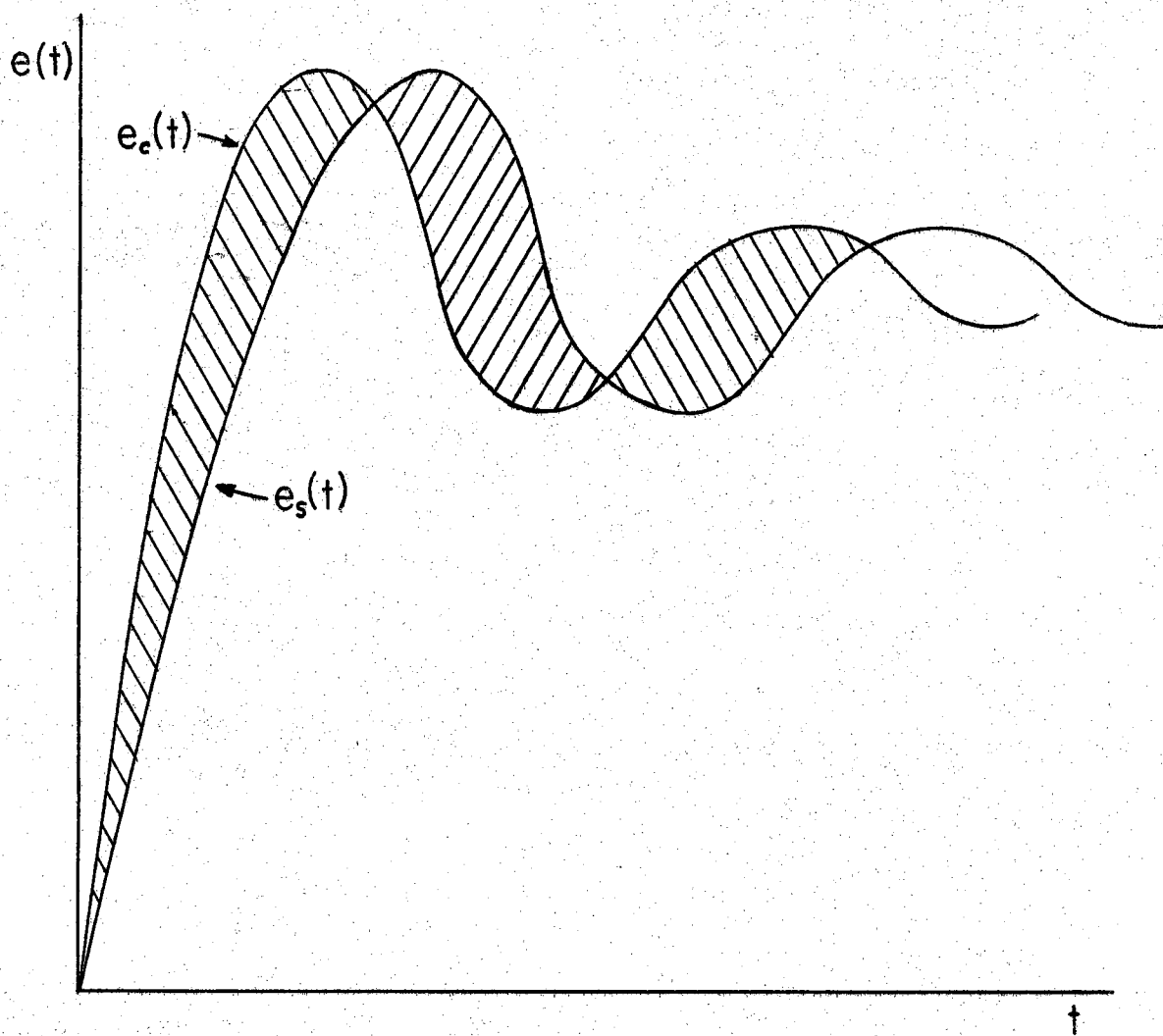
Figure A-3

Possible System Responses

is considered as good a yardstick as any other.

With reference to Figure A.1

$$G(s)\, G_{HO}(s) = \frac{K}{s(1+sT_a)}\, \frac{1-\epsilon^{-Ts}}{s} \tag{A.7}$$

using the z-transform one can obtain:

$$\overline{GG}_{HO}(z) = K\left[\frac{T}{z-1} - T_a + \frac{(z-1)\,T_a}{z-\epsilon^{-T/T_a}}\right] \tag{A.8}$$

Therefore

$$\frac{C(z)}{R(z)} = \frac{\overline{GG}_{HO}(z)}{1 + \overline{GG}_{HO}(z)} \tag{A.9}$$

and

$$\frac{C(z)}{R(z)} = \frac{e\,(z+f)}{z^2+gz+h} \tag{A.10}$$

where e, f, g and h are functions of the basic variables K, T_a and T and are constant for a particular choice of these variables.

The characteristic equation that results from a particular choice of the basic constants must now be checked for stability, for, though a second order continuous system can never be unstable for positive constants, it is possible for the same system, when sampled, to be unstable.

The solution at the output terminals for the sampled system using z-transform analysis consists of a sequence of impulses c(nT). The continuous time function c(t) that results from the inverse z-transform operation on C(z) may be used as a good approximation to the analog output [17], as it automatically joins the points indicated by the impulse sequence.

With the step input used to derive the output from the continuous system, equation (A.2), i.e., $R(s) = \frac{A}{s}$ or $R(z) = \frac{Az}{z-1}$, the continuous time function derived from a z-transform analysis is:

$$c_s(t) = A + F\,\epsilon^{-Gt}\, \cos\,(Rt + \Psi). \tag{A.11}$$

here P, Q, R and Ψ are again functions of the basic variables K, T_a and T. Rewriting equation (A.2) as

$$c_c(t) = A - B\varepsilon^{-Ct} \sin(Dt + \phi) \tag{A.12}$$

B, C, D and ϕ functions of K, T_a and T, and substituting (A.11) and (A.12) into (A.3) the performance index becomes:

$$P.I. = \int_0^{\infty} \left[-B\varepsilon^{-Ct} \sin(Dt+\phi) - P\varepsilon^{-Qt}\cos(Rt+\Psi) \right]^2 dt, \tag{A.13}$$

which is clearly integrable.

A digital computer program was written to instrument the algebraic operations necessary to check stability, determine the constants associated with the analog outputs of both systems, and determine the integral (A.13) as a number. The data taken from the computer runs are plotted as a function of the sampling period in Figures A.4, A.5 and A.6. The basic variables K and T_a were iterated, as indicated in Figure A.7 to maintain a fixed damping ratio, $\int$, for different system natural frequencies.

In order to interpret these graphs it is first necessary to examine equation (A.5) an analytic expression for the performance index. The Laplace Transform of the continuous system output can be written from equation A.1 as:

$$C_c(s) = \frac{AK/T_a}{s(s + \frac{1}{T_a}s + \frac{K}{T_a})}. \tag{A.14}$$

The Laplace Transform of the time function representing the output of the sampled system, i.e. the smooth curve through the sample points, can be derived from equation (A.10) using the relation $z = \varepsilon^{Ts}$

$$C(s) = \frac{\varepsilon^{Ts}Ae(\varepsilon^{Ts}+f)}{(\varepsilon^{Ts}-1)(\varepsilon^{2Ts}+g\varepsilon^{Ts}+h)}. \tag{A.15}$$
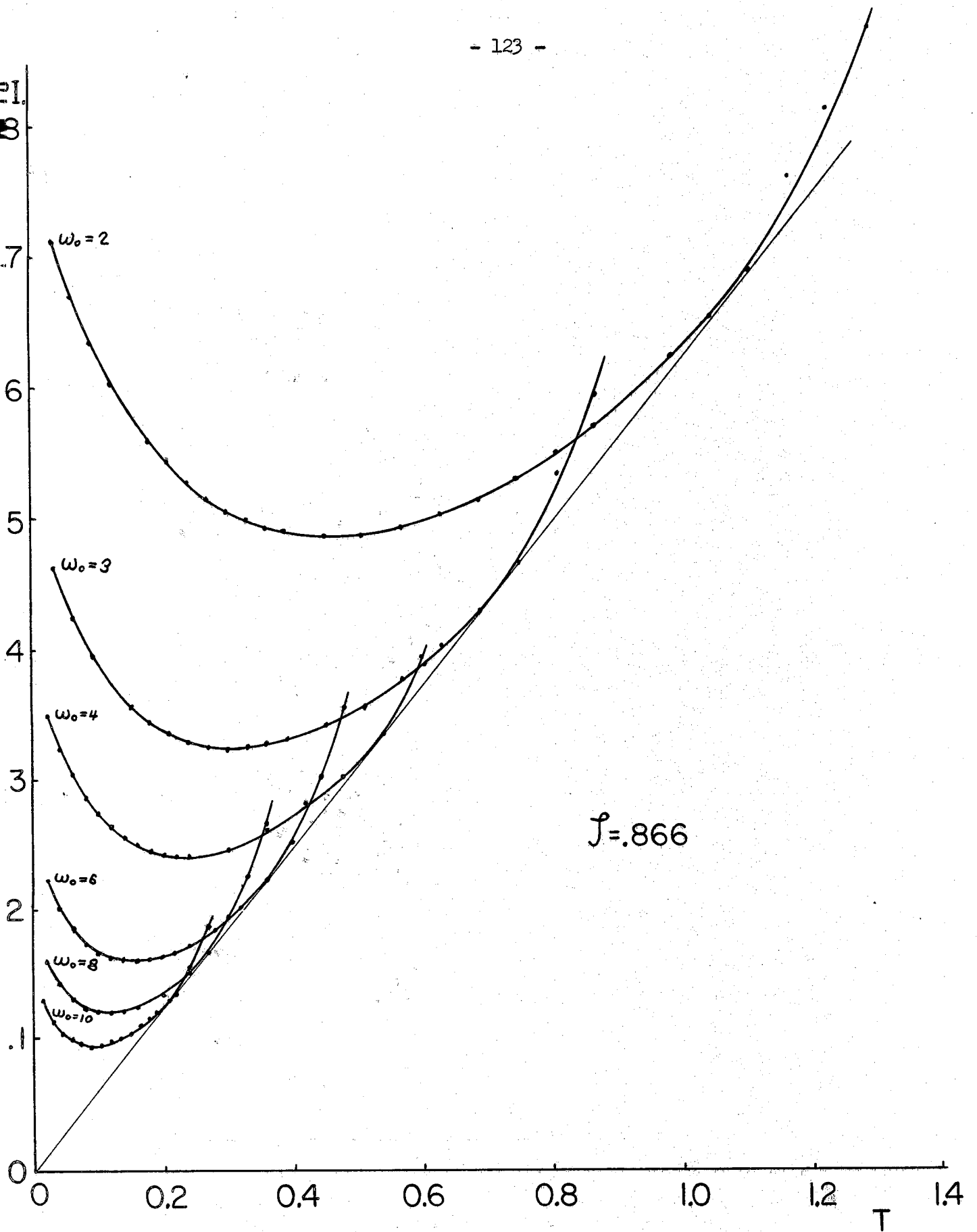
Figure A-4

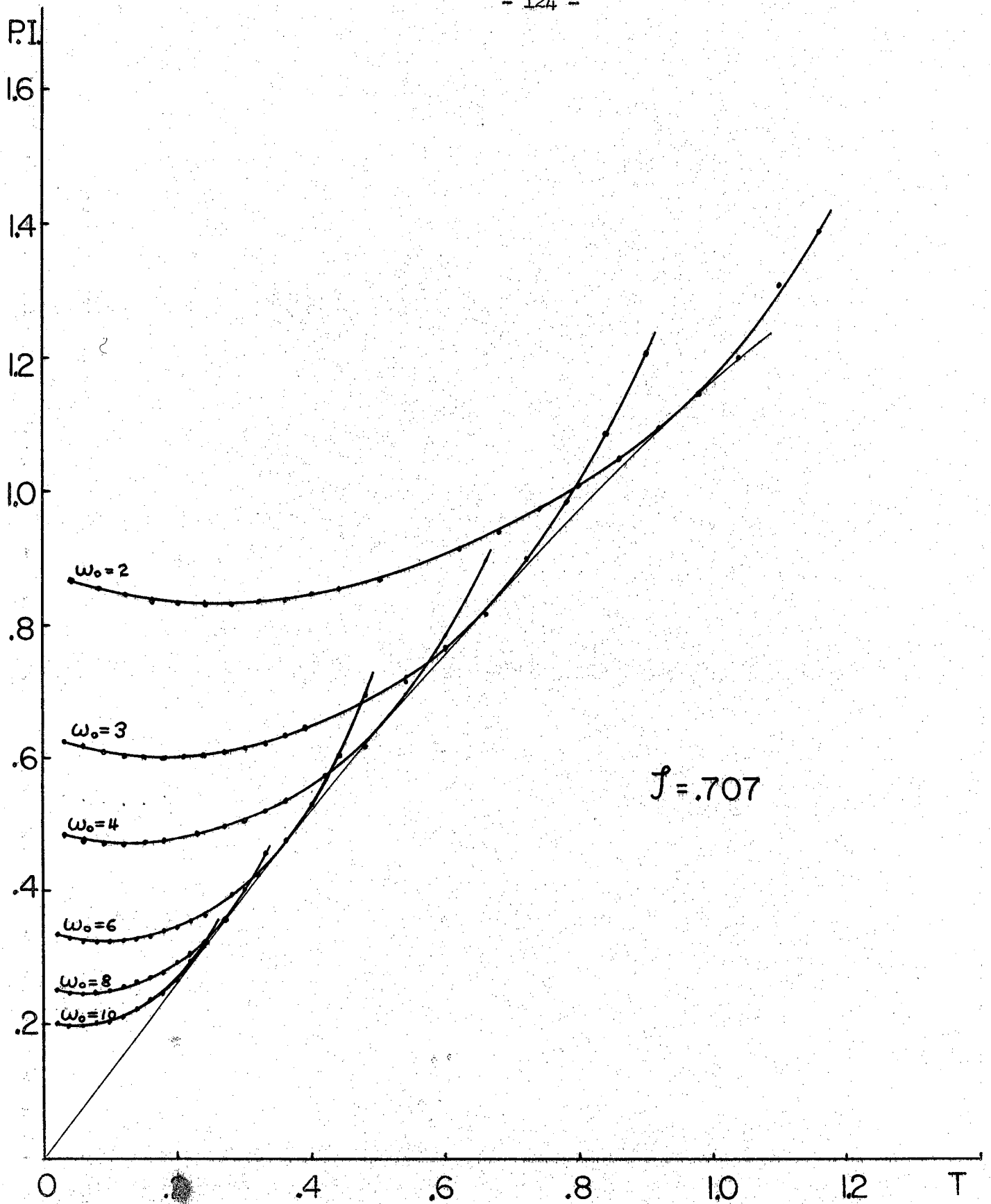Plot of Performance Index Against Sampling Period

Figure A-5

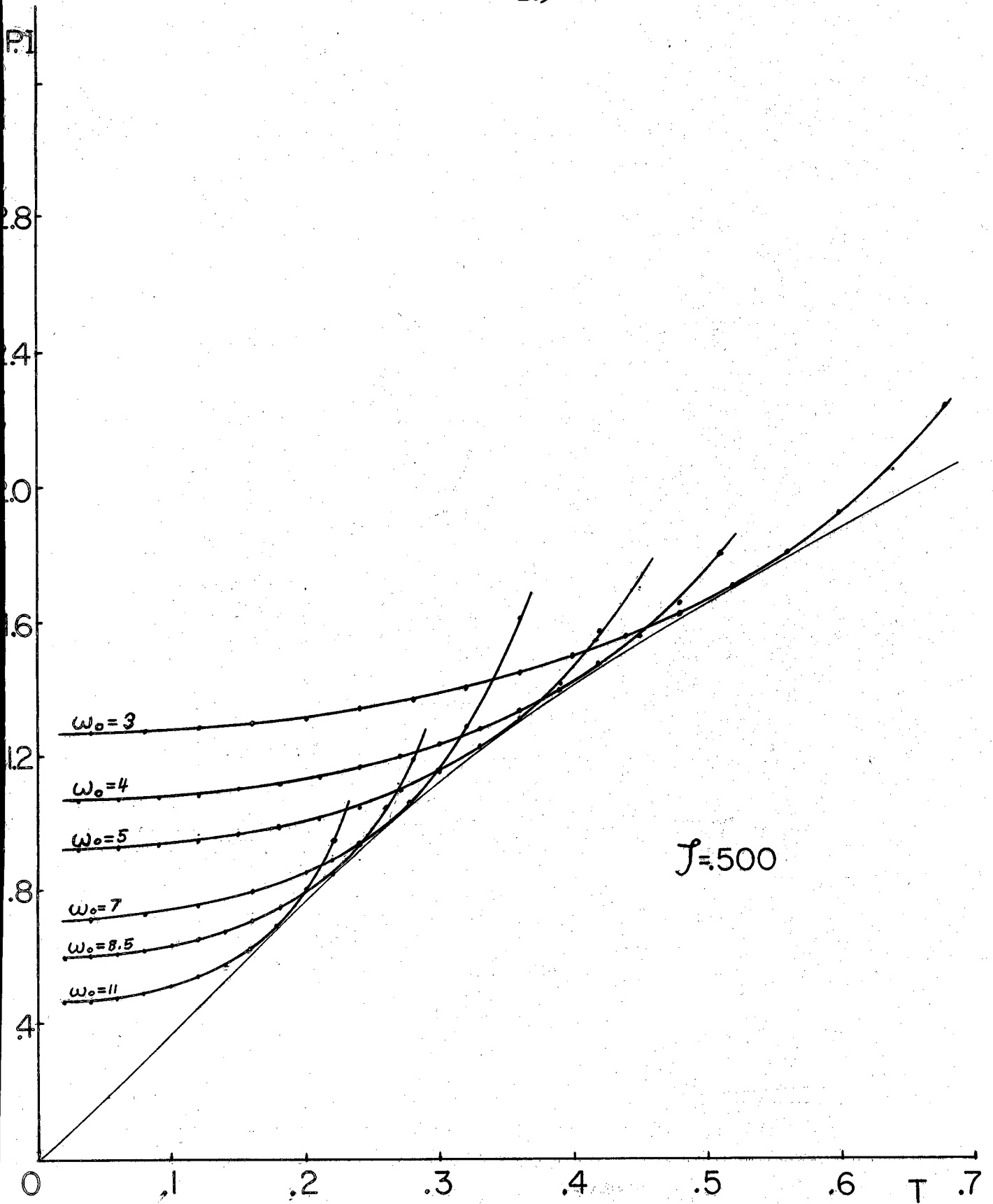Plot of Performance Index Against Sampling Period

Figure A-6

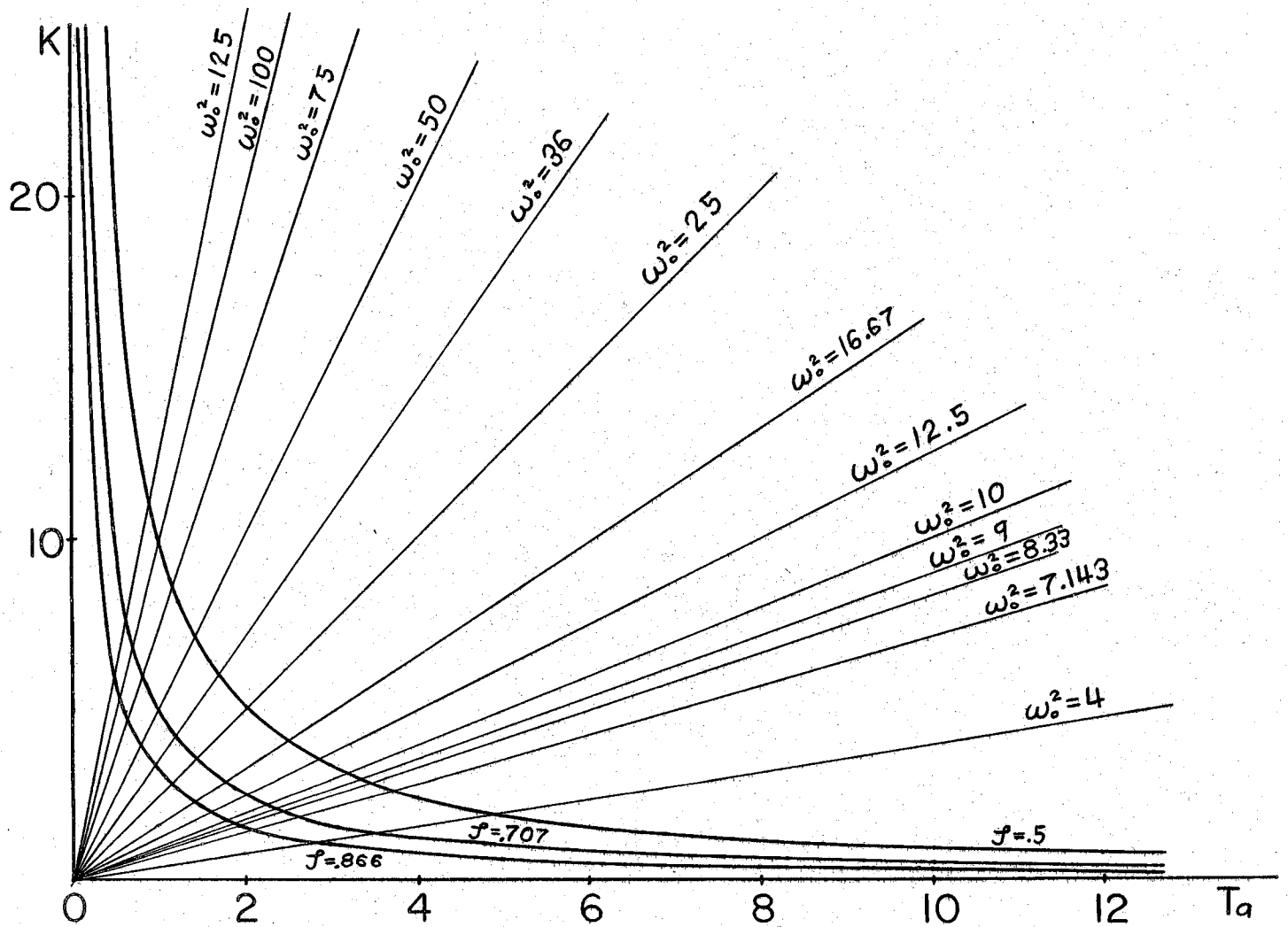Plot of Performance Index Against Sampling Period

Figure A-7

Second Order System Parameters

Consequently

$$C(s) = \left[\frac{AK/T_a}{s(s + \frac{1}{T_a}s + \frac{K}{T_a})} - \frac{\varepsilon^{Ts}Ae(\varepsilon^{Ts}+f)}{(\varepsilon^{Ts}-1)(\varepsilon^{2Ts}+g\varepsilon^{Ts}+h)} \right], \qquad (A.16)$$

and therefore

$$P.I. = \frac{1}{2\pi j} \int_{-\infty}^{\infty} \left\{ \left[\frac{AK/T_a}{s(s^2+\frac{1}{T_a}s+\frac{K}{T_a})} - \frac{\varepsilon^{Ts}Ae(\varepsilon^{Ts}+f)}{(\varepsilon^{Ts}-1)(\varepsilon^{2Ts}+g\varepsilon^{Ts}+h)} \right] \cdot \right.$$

$$\left. \left[\frac{Aef\varepsilon^{Ts}(\varepsilon^{Ts}+\frac{1}{f})}{h(1-\varepsilon^{Ts})(\varepsilon^{2Ts}+\frac{g}{h}\varepsilon^{Ts}+\frac{1}{h})} - \frac{AK/T_a}{s(s^2-\frac{1}{Ta}s+\frac{K}{T_a})} \right] \right\} ds \qquad (A.17)$$

Examination of the integrand in equation (A.17) reveals that this
quantity approaches infinity as T approaches zero. This result is
directly due to the effect of the hold circuit in equation (A.17) and
can be checked readily. The validity of this limiting process is in
doubt, however, due to the method of modulation used in connection
with the sampling device ([1], p. 568); but, as attention will not be
focussed in this region (T→0) the method is acceptable.

Examination of the integrand reveals also that the quantity vanishes
as the complex variable of integration approaches infinity. The integral
can thus be considered as a contour integral with the path of integration
choosen to include all poles of the integrand function. A suitable
contour in the complex plane will be the imaginary axis together with a
semi-circle of infinite radius in the left half-plane, or in the right
half-plane, as is appropriate to the pole under consideration. The
section of this path at infinity will not contribute to the integral,
thus leaving the value of the contour integral equal to that of the
original line integral, equation (A.17). The contour integral can now

be evaluated from residue theory:

$$P.I. = \sum \text{residue of} \left[C_c(s) \ C_e(-s) \right] \qquad (A.18)$$

The two parts of the integrand of equation (A.17) can be expressed in factored form and then multiplied out. The partial fractions that result are quite numerous (many will have an infinite number of poles due to the exponential terms present) and the calculation of the residues is, in all but a few cases, a complicated procedure. This analytical approach will not be pursued in this report, as the yield would be the equation of the curves already plotted. The characteristics of these curves, for the second order system choosen, can be seen satisfactorily from the graphs plotted.

The graph of Figure A.4 shows that, as might be expected, the performance index increases with increase in sampling period. It further shows that plots of P.I. versus T are confined within a region of the P.I. - T space with the region boundary defined by a curve tangent to the plots. In Figure A.4 the boundary curve appears to be a straight line. The Figures A.5 and A.6 have exactly similar characteristics, but in these cases the boundary curves are lines with a large radius of curvature.

The question of performance is now raised, and a sampling period is sought below which the sampled system is "nearly" equal in performance to the continuous system, and above which this is not so. It is apparent that any such point must be somewhat arbitary as the system characteristics cannot change in any manner resembling a step function.

The sampling period corresponding to the point of tangency with the boundary curve is clearly a choice for the point sought. For sampling

periods smaller than that corresponding to the point of tangency, the rate of change of PI with respect to T is small. Above the point of tangency the rate of change becomes larger, rapidly.

The sampling frequency associated with each point of tangency is plotted against the system bandwidth, $BW = \omega_o \left[1-2\zeta^2 + \sqrt{4\zeta^4 - 4\zeta^2 + 2} \right]^{\frac{1}{2}}$, of the system in question in Figure A.8. The points plotted define straight lines for each different system damping ratio, and the slope of these lines increases with the magnitude of the damping ratio. Along any one line the relation between sampling frequency and bandwidth is:

$$\omega_s = K \cdot BW \tag{A.19}$$

where K is a function of the system.

This result indicates that for the performance of a sampled data system to be equal "nearly" to that of the identical continuous system the sampling frequency must be at least K times as great as the system bandwidth. Acknowledging that this investigation has been concerned with second order systems and that a system bandwidth is related only empirically to the greatest input frequency anticipated, this result cannot be interpreted literally with systems of any order. Figure A.8 shows, however, that a relationship exists between the sampling frequency and the system bandwidth (and consequently the input frequency) and that a "turning point" exists.

Systems that can be considered approximately second order are designed frequently for a ζ of the order of 0.7. The graph of Figure A.8 indicates that K should be of the order of 3. To allow for variations of the design criteria and to allow for systems of higher order a
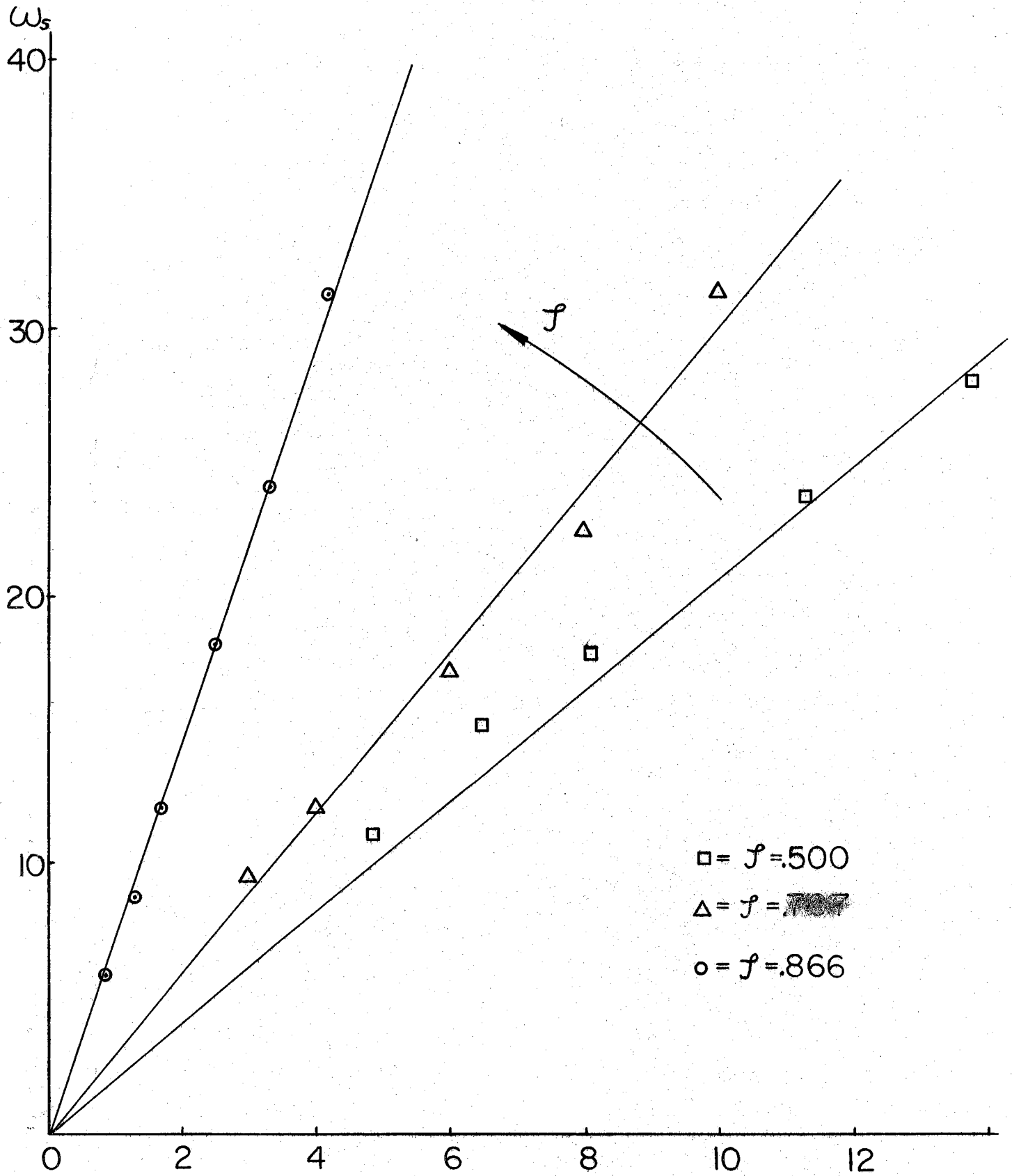
Figure A-8

Sampling Frequency Plotted Against In-put Frequency

"factor of safety" is incorporated thus increasing the value of K.

It is recommended that this factor of safety be of the order of 3 to 4, and thus equation (A.19) is rewritten as:

$$\omega_s = 10\omega_i \text{ or } 10 \text{ BW.} \hspace{3cm} \text{(A.20)}$$

APPENDIX B

Abbreviated Translation of the Section

on Performance Indices from The Book

by

M. A. Aizerman

LECTURES ON THE THEORY

OF AUTOMATIC CONTROL

Second Edition

Moscow, 1958

a)   General Remarks

In the preceding analysis the values of coordinates of the response with respect to zero, i.e., with respect to the original equilibrium state of the control process were calculated. The response of the system was caused by the initial action of disturbances. In this paragraph we are interested in calculating the changes of the values of coordinates with respect to a new equilibrium state, occurring in the system as a result of existing disturbances. Only unit step function disturbances, 1, will be considered. This limitation is, however, not essential for the application of the integral performance indices. Except for computational complications they can be easily extended to different functions of disturbances. In restricting the investigation to the performance indices of transient responses, we will replace the unit step disturbances 1 by their equivalent initial conditions.

A transient response would be ideal if, at the instant of the application of unit step disturbance, the coordinate under consideration

would instantaneously take on its new steady state value and remain at this value until the application of a new disturbance (Fig. 165). In actual systems such response is impossible. However, the smaller the area (shaded in Fig. 166) between the actual and ideal responses, the less does the actual response differ from the ideal response. If there is no overshoot (if the system is position control, see Fig. 167) or if the curve $\bar{x}(t)$ does not repeatedly intersect the time axis t (if the system is not position control) this area is defined by the integral

$$\int_0^\infty x\,(t)\,dt. \tag{4.30}$$

In other cases the above integral does not define the above considered area, since in the evaluation of the integral the consecutive areas are added up with opposite signs (Fig. 168).

Thus, for example, in the case of slowly decaying oscillations the integral would be small, regardless of the amplitudes, while the area describing the deviation between the actual and the ideal response may be arbitrarily large.

In the above discussed cases, when the integral (4.30) defines the given area, it serves as a convenient means to select the system para- meters. The parameters are selected in such a way as to minimize this integral. It is obvious that such a performance index is indirect (unreliable, Z.V.R.) and can only be used for preliminary selection of parameters, since it admits perfect oscillations with equal positive and negative areas in the response. Nevertheless such a performance index frequently enables one to make a rapid initial estimate of the system parameters. The validity of such selection of parameters can be proved
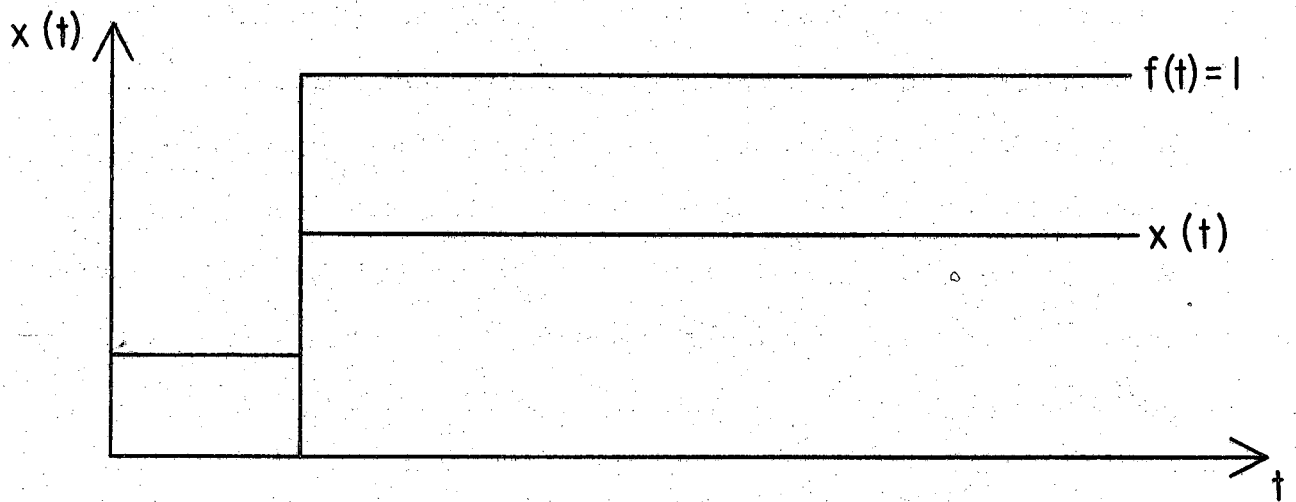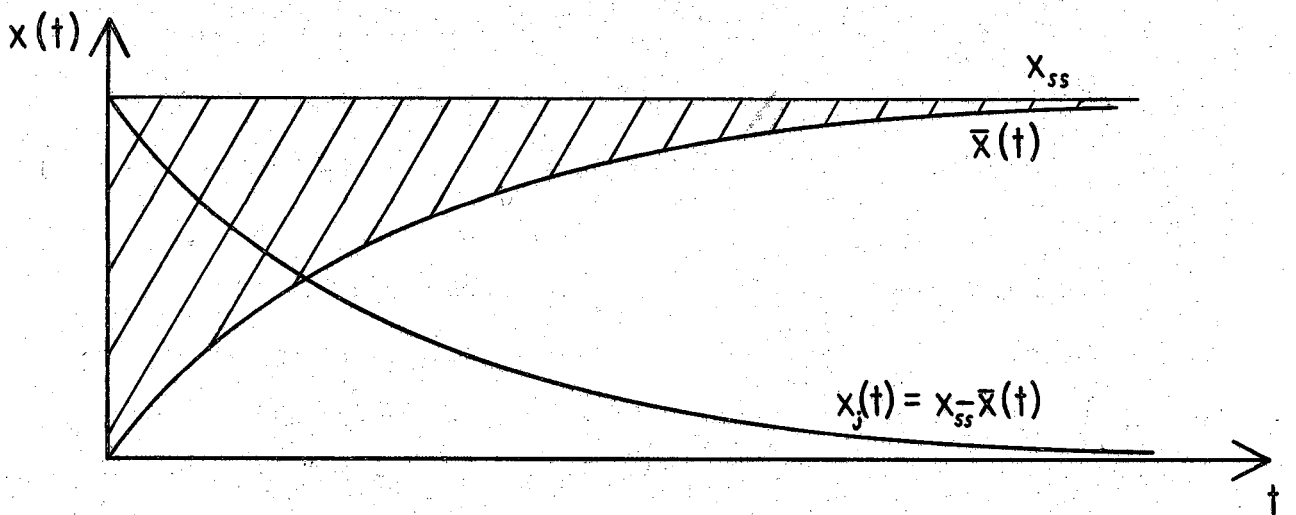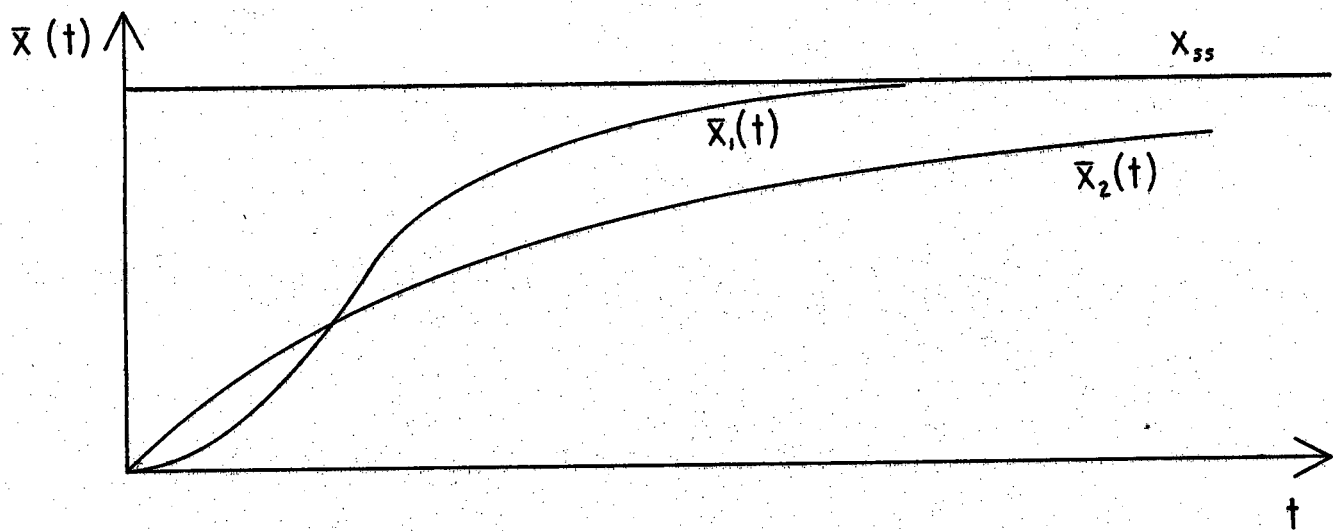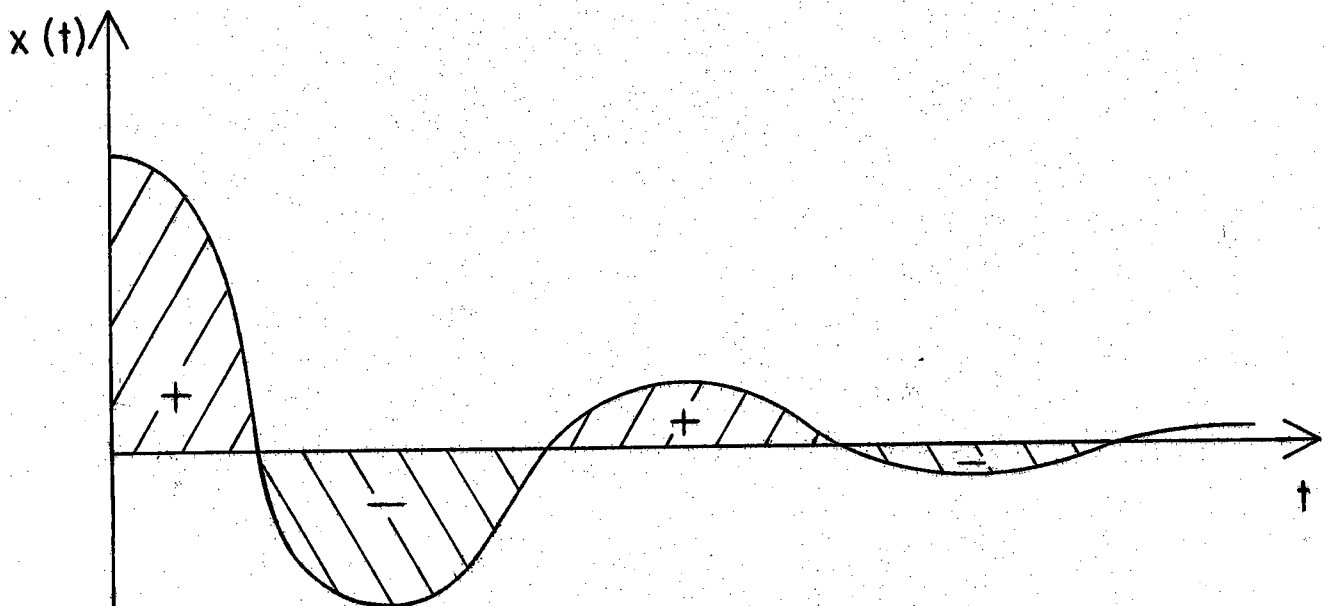
Figure 165

Figure 166

Figure 167

Figure 168

later from the recording of the response.

To evaluate the integral (4.30) let us note that the Laplace transform of the function $x_i(t)$ is, by definition

$$\mathcal{L}[x_i(t)] = \int_0^\infty x_i(t)\epsilon^{-pt}\, dt$$

and consequently

$$\int_0^\infty x_i(t)\, dt = \lim_{p \to 0} \mathcal{L}[x(t)] .$$

The practical application of such a simple performance index is not feasable, since it is seldom apparent in advance that the response does not overshoot or that in non-positional systems the controlled coordinate does not reach zero value several times during the course of the response.

If the response is oscillatory, the proximity of the transient response to the ideal one may be estimated from the integral $\int_0^\infty |x(t)|\, dt$; this integral is, however, difficult to compute. It is more convenient to use the integral

$$\int_0^\infty x^2(t)\, dt \tag{4.31}$$

as a performance index of the response. If the system parameters are selected by minimizing this integral, the transient response thus obtained is usually excessively oscillatory.

In order to avoid too oscillatory responses, it was proposed to select the system parameters by minimizing the integral

$$\int_0^\infty \left[x^2(t) + \tau^2\, \dot{x}^2(t)\right] dt \tag{4.32}$$

where τ — a real arbitrary constant.

The selection of system parameters by minimizing the value of this integral (if one can decide upon the value of the constant τ) yields sufficiently good transient response with small overshoot, frequently even a monotonic response. Occasionally one makes use of a more complex performance index

$$\int_0^\infty \left[x^2(t) + \tau_1^2 \dot{x}^2(t) + \tau_2^4 \ddot{x}^2(t) \right] \, dt$$

or, in a more general form.

$$\int_0^\infty \left(x^2(t) + \tau_1^2 \left[\frac{dx(t)}{dt} \right]^2 + \tau_2^4 \left[\frac{d^2 x(t)}{dt^2} \right]^2 + \dots \right.$$
$$\left. \tau_n^{2n} \left[\frac{d^n x(t)}{dt^n} \right]^2 \right) \, dt$$

We will restrict ourselves to the simpliest performance index (4.32). In order to make use of this integral performance index in the design of systems the following questions have to be answered:

1. How does one select the constant τ in (4.32) in the analysis of actual automatic control systems?

2. How does one find the parameters of the system such that the selected performance index is minimized?

3. How close will the transient response, obtained by selecting the parameters in this fashion, approach the response which best satisfies the specifications?

b) Selection of the Integral Performance Index

Let us write the integral

$$I = \int_0^\infty \left[x^2(t) + \tau^2 \dot{x}^2(t) \right] \, dt$$

in the form of the difference of two integrals:

$$\int_0^\infty x^2(t) + \tau^2 \dot{x}^2(t) \quad dt =$$

$$= \int_0^\infty \left[x(t) + \tau \dot{x}(t)\right]^2 dt - 2\tau \int_0^\infty \dot{x}(t) \, x(t) \, dt =$$

$$= \int_0^\infty \left[x(t) + \tau \dot{x}(t)\right]^2 dt - 2\tau \int_0^\infty x(t) \, \frac{dx}{dt} \, dt =$$

$$= \int_0^\infty \left[x(t) + \tau \dot{x}(t)\right]^2 dt - 2\tau \int_0^\infty x(t) \, dx$$

Let us evaluate the last integral

$$2\tau \int_0^\infty x(t) \, dx = 2\tau \left. \frac{x^2(t)}{2} \right]_0^\infty = \tau \left[x^2(\infty) - x^2(o)\right].$$

If the system is stable then $x(\infty) = 0$, since it was assumed at the beginning of this paragraph that the value of x approaches x_{ss} as $t \rightarrow \infty$. Then

$$\int_0^\infty \left[x^2(t) + \tau^2 \dot{x}^2(t)\right] dt = \int_0^\infty \left[x(t) + \tau \dot{x}(t)\right]^2 dt + \tau x^2(o).$$

The last term on the right side is a constant quantity determined by the initial conditions on the system. The original integral

$$\int_0^\infty \left[x^2(t) + \tau^2 \dot{x}^2(t)\right] dt$$

will take on its minimum value if the integral on the right side of the previous equation approaches zero,

$$\int_0^\infty \left[x(t) + \tau \dot{x}(t)\right]^2 dt = 0.$$

Since the integrand is always positive, this can be satisfied only if the integrand function is equal to zero, i.e., if

$$\left[x(t) + \tau \, \dot{x}(t) \right]^2 = 0$$

or

$$x(t) + \tau \, \dot{x}(t) = 0 \qquad\qquad (4.33)$$

Hence the original integral is a minimum if $x(t)$ satisfies the differential equation (4.33). Its lowest minimum value $I_{min\ min}$ is equal to

$$I_{min\ min} = x^2(o).$$

The differential equation (4.33) defines the transient response which can be approached in the limit, if it is possible to select the parameters in such a way that $I = I_{min\ min}$. This optimum response is described by the exponential $x(t) = x(o)\epsilon^{-\frac{t}{\tau}}$.

The _value_ _of_ τ _shall_ _be_ _selected_ _in_ _such_ _a_ _way_ _that_ _the_ _exponential_ $x(t) = x(o)\epsilon^{-\frac{t}{\tau}}$ _will_ _satisfy_ _the_ _specifications_ _of_ _the_ _transient_ _response._ The selection of the constant τ fixes the integral performance index. Henceforth the system parameters are selected in such a way as to minimize the adopted integral performance index, I_{min}. The numerical values of I_{min} obtained in every actual case are obviously greater than $I_{min\ min}$, and the system response will differ from the indicated exponential. Of all possible system parameters, however, the parameters determined in this way will yield the response which is closest to the exponential.

It can be shown that the minimization of a simplier quadratic integral performance index $\int_{0}^{\infty} x(t)\,dt$ guarantees that the response will approach

$$x(t) = x(o) \frac{\sin \omega_c t}{\omega_c t} \ .$$

The plot of this function is shown in Fig. 169.

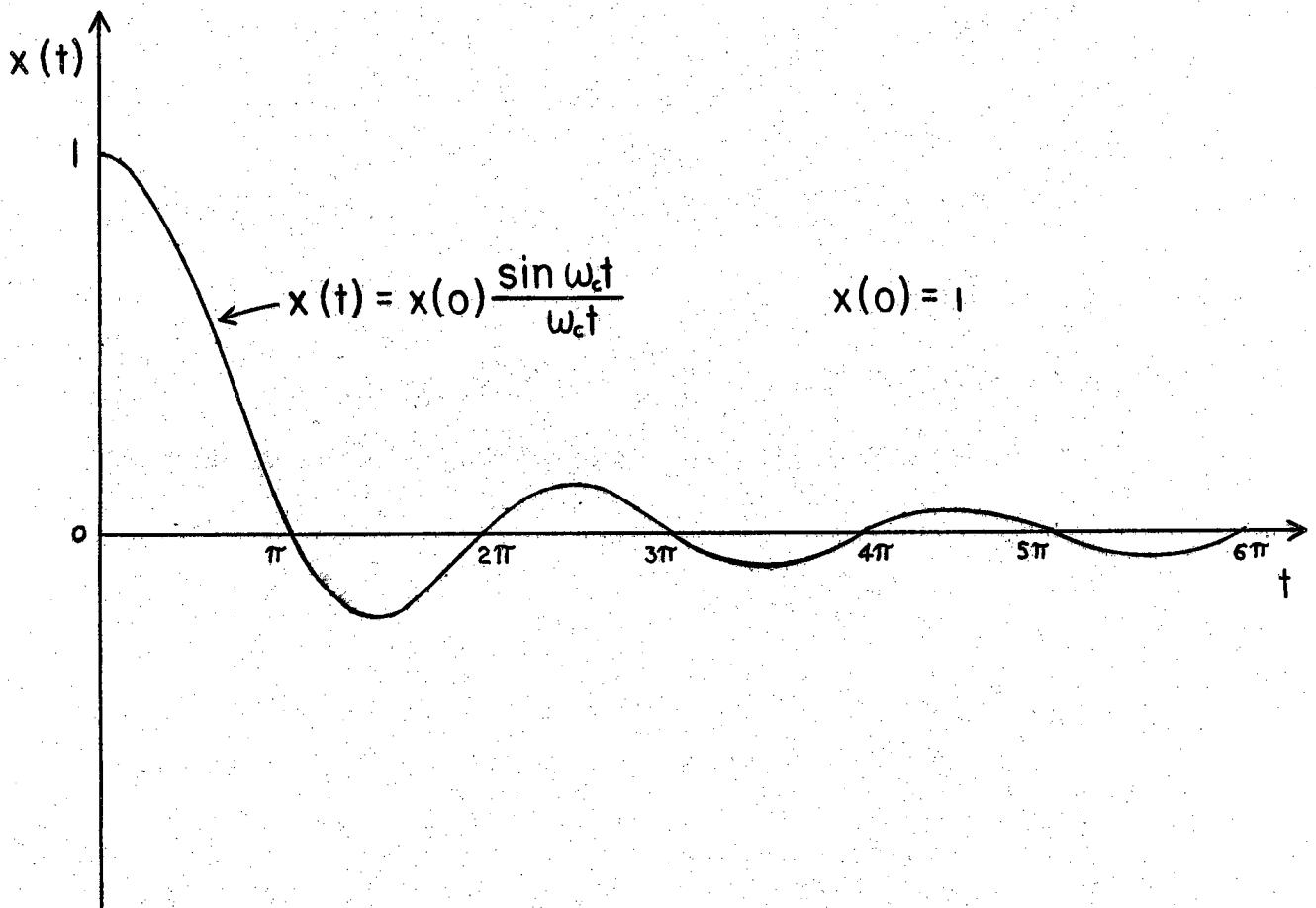$$x(t) = x(o) \frac{\sin \omega_c t}{\omega_c t} \qquad x(o) = 1$$

Figure 169

Hence it is apparent that the selection of system parameters by minimization of the integral $\int_0^\infty x^2 dt$ yields systems with excessive oscillations in the transient response.

By making use of the more complex performance indices, it is possible, by minimizing them, to approach the responses of a more complex form, for example, the response consisting of several exponential terms.

Let us consider now the evaluation of $\int_0^\infty (x^2 + \tau^2 \dot{x}^2)\, dt$ in terms of the system parameters and the selection of system parameters to minimize this integral.

c) Determination of System Parameters which Minimize the Integral Performance Index.

Consider the system of linear differential equations of the general form*

$$\dot{x}_i = \sum_{j=1}^{n} a_{ij} x_j$$

and the most general form of the quadratic performance index

$$I = \int_0^\infty V\, dt,$$

where

$$V = A_1 x_1^2 + A_2 x_2^2 + \ldots + A_n x_n^2 = \sum_{i=1}^{n} A_i x_i^2$$

The particular performance index (4.32) which is of interest here is

---

* If the system is of order higher than one it is easy to arrange it in the above form by designating the derivatives of the original variables as new variables. E.g., the equation $a\ddot{x}_1 + b\dot{x}_1 + cx_1 = 0$ is rearranged as

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = -\frac{b}{a} x_2 - \frac{c}{a} x_1$$

obtained by letting $x_1 = x$, $x_2 = \dot{x}_1$, $A_1 = 1$, $A_2 = \tau^2$ and $A_3 = A_4 = \ldots = A_n = 0$.

Let us select another quadratic form U such that

$$\frac{dU}{dt} = -V. \hspace{3cm} (4.34)$$

Then it is easy to calculate the integral $I = \int_0^\infty V\, dt$. Obviously

$$V dt = -dU$$

i.e.,

$$I = \int_0^\infty V dt = -\int_0^\infty dU = -U\Big]_0^\infty = -\left[U(\infty) - U(o)\right]$$

In a stable system $x_1 = x_2 = \ldots = x_n = 0$ at $t = \infty$ and hence $U(\infty) = 0$.

Thus

$$I = \int_0^\infty V\, dt = U(o)$$

i.e., in order to evaluate I it is necessary to substitute into U the initial values of x_1, x_2, $\ldots x_n$.

In order to evaluate U we will assume it to be of the type

$$U = \sum_{i,\, j=1}^{n} B_{ij} x_i x_j$$

where all B_{ij}'s are the numbers which have to be selected in such a way as to satisfy the equation

$$\frac{dU}{dt} = -V$$

or

$$\sum_{i=1}^{n} \frac{\partial U}{\partial x_i} \dot{x}_i = -V$$

Substituting into this the expressions for U and V we get:

$$\sum_{i=1}^{n} \left[\sum_{j=1}^{n} B_{ij} x_j \right] \dot{x}_i = - \sum_{i=1}^{n} A_i x_1^2.$$

Substitution for x_i the original, linear first order system differ-

ential equations yields

$$\sum_{i=1}^{n} \left[\sum_{j=1}^{n} B_{ij} x_j \right] \left[\sum_{j=1}^{n} a_{ij} x_j \right] = - \sum_{i=1}^{n} A_i x_i^2 .$$

Both the left and the right-hand side of the above equation consist of quadratic forms. By equating the coefficients of $x_1^2, x_2^2, \ldots x_n^2$ of both sides and by setting the coefficients of the $x_i x_j$ - terms $(i \neq j)$ equal to zero (since there are no such terms on the right side of the equation), we get a set of linear algebraic equations containing all B_{ij}'s. The solution of these equations yields the coefficients B_{ij} of U, and then, if the initial conditions are given, one finds U as a function of system parameters (i.e., of the coefficients a_{ij}) which is then subject to minimization.

Example 1. We illustrate the method of finding U for the example of equation

$$a_o \ddot{x}_1 + a_1 \dot{x}_1 + a_2 x_1 = 0,$$

which we arrange as

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = \frac{a_1}{a_o} x_2 - \frac{a_2}{a_o} x_1$$

Let

$$V = x_1^2 + \tau^2 x_2^2 .$$

We are trying to find U of the type

$$U = B_{x_1} x_1^2 + B_{x_1 x_2} x_1 x_2 + B_{x_2} x_2^2$$

From the relationship

$$\frac{dU}{dt} = -V$$

we find

$$\frac{\partial U}{\partial x_1} \dot{x}_1 + \frac{\partial U}{\partial x_2} \dot{x}_2 = (2B_{x_1} x_1 + B_{x_1 x_2} x_2) x_2 +$$

$$+ (2B_{x_2} x_2 + B_{x_1 x_2} x_1) \left(-\frac{a_1}{a_0} x_2 - \frac{a_2}{a_0} x_1 \right) =$$

$$= - x_1^2 - \tau^2 x_2^2 .$$

Equating the coefficients on both sides we get the system of equations

$$- \frac{a_2}{a_0} B_{x_1 x_2} = - 1$$

$$B_{x_1 x_2} - 2 \frac{a_1}{a_0} B_{x_2} = -$$

$$2 B_{x_1} - \frac{a_1}{a_0} B_{x_1 x_2} = 2 \frac{a_2}{a_0} B_{x_2} = 0.$$

The solution of the above system of equations yields

$$B_{x_1 x_2} = \frac{a_0}{a_2}$$

$$B_{x_1} = \frac{a_0^2}{2a_1 a_2} \quad \frac{a_2^2}{a_0^2} 2 + \frac{a_1^2}{a_0^2} + \frac{a_2}{a_0}$$

$$B_{x_2} = \frac{a_0}{2a_1} \quad \frac{a_0}{a_2} + 2$$

from which

$$I = \int_0^\infty (x_1^2 + \tau^2 \dot{x}_1^2) \, dt = U(0) =$$

$$= \frac{a_0^2}{2a_1 a_2} \left[\frac{a_2^2}{a_0^2} \tau^2 + \left(\frac{a_1^2}{a_0^2} + \frac{a_2}{a_0} \right) \right] x_1^2(0) +$$

$$+ \frac{a_0}{2a_1} \left[\frac{a_0}{a_2} + \tau^2 \right] x_2^2(0) + \frac{a_0}{a_2} x_1(0) x_2(0).$$

The values of parameters a_0, a_1, and a_2 for minimum I can now be found easily by the general rules for finding the absolute minimum of a function of several arguments.

Example 2. Let the system under consideration be described by the second

order equation

$$\frac{d^2x}{dt^2} + h \frac{dx}{dt} + 3x = 0$$

Let, at $t = o$, $x = 1$, $\frac{dx}{dt} = 0$. The integral performance index is

$I = \int_0^\infty (x^2 + \tau^2 \dot{x}^2) dt$. In the case under consideration

$$x_1 = x \qquad x_2 = \frac{dx}{dt}, \; x_1(o) = 1, \; x_2(o) = o,$$

$$\frac{a_1}{a_o} = h, \; \frac{a_2}{a_o} = 3, \; \tau^2 = 2.$$

Hence

$$I = \frac{1}{6h} \left[(9)(2) + (3 + h^2) \right] = \frac{21 + h^2}{6h}$$

The problem requires to find an $h = h_{min}$ at which $I = I_{min}$. To accomplish

this we set the derivative $\frac{dI}{dh}$ equal to zero.

$$\frac{dI}{dh} = \frac{1}{6} - \frac{7}{2h^2} = 0 ;$$

This relationship yields the quantity h_{min} as $h_{min} = \sqrt{21}$ and conse-

$$I_{min} = B_x^* = \frac{\sqrt{21}}{6} + \frac{7}{2 \sqrt{21}} = 1.53.$$

In this case

$$I_{min \; min} = \sqrt{2} \approx 1.41.$$

Consequently, when $h = \sqrt{21}$, the response falls closest to exponential

with the exponent of $-\frac{1}{\sqrt{2}} t$; however it still differs from this ideal

response.

d) <u>Estimate of the Deviation from the Optimum Response</u>.

In the solution of practical problems, it is frequently desirable to

estimate the deviation of the actual system response from the response to

which one strives by minimizing the integral performance index.

Let the value of the integral, I_{min}, be evaluated for the selected values of system parameters.

Also known is the value, $I_{min\ min}$, of this integral performance index, at which the transient response coincides with the exponential $x^* = C \in -\frac{t}{\tau}$.

It has been shown above that the value of $I = I_{min\ min}$ is determined by the square of the initial deviation $x^2(o)$ and τ :

$$I_{min\ min} = \tau x^2(o).$$

Let us investigate the difference between the two integral performance indices

$$\mathcal{E} = I_{min} - I_{min\ min}$$

Let $\triangle x = x - x^*$, where x is the variation of the coordinate under consideration for the chosen values of parameters and x^* - variation of the same coordinate at $I = I_{min}$.

We substitute the new variables into the general expression for I_{min}:

$$I_{min} = \int_o^\infty \left[(x^* + \triangle x)^2 + \tau^2 \ \left(\frac{dx^*}{dt} \ \frac{d\ \triangle x}{dt} \right)^2 \right] dt,$$

or

$$I_{min} = \int_o^\infty \left(\left[x^{*2} + \tau^2 \left(\frac{dx^*}{dt} \right)^2 \right] + \left[\triangle x^2 + \tau^2 \left(\frac{d\triangle x}{dt} \right)^2 \right] \right.$$

$$\left. + 2 \left(x^* \triangle x + \tau^2 \frac{dx^*}{dt} \frac{d \triangle x}{dt} \right) \right) dt =$$

$$= \int_o^\infty \left[x^{*2} + \tau^2 \left(\frac{dx^*}{dt} \right)^2 \right] dt + \int_o^\infty \left[\triangle x^2 + \tau^2 \left(\frac{d \triangle x}{dt} \right)^2 \right] dt +$$

$$+ 2 \int_o^\infty \left(x^* \ \triangle x + \tau^2 \frac{dx^*}{dt} \frac{d\triangle x}{dt} \right) dt$$

By substituting

$$x^* = C \in \frac{-t}{\tau}$$

into the above integral and integrating by parts it can be shown that the last integral is equal to zero.

The first integral is equal to $I_{min\ min}$. Consequently

$$I_{min} = I_{min\ min} + \int_0^\infty \left[\Delta x^2 + \tau^2 \left(\frac{d\ \Delta x}{dt}\right)^2 \right] dt$$

Transposing $I_{min\ min}$ to the left hand side we get

$$\mathcal{E} = I_{min} - I_{min\ min} = \int_0^\infty \left[\Delta x^2 + \tau^2 \left(\frac{d\ \Delta x}{dt}\right)^2 \right] dt.$$

Consequently the difference of the two integral performance indices is defined by the same integral; only the variables x and $\frac{dx}{dt}$ have to be replaced by Δx and $\frac{d\ \Delta x}{dt}$. In the further evaluation we will make use of the well-known Buniakovski's inequality:

$$\int_a^b f_1(t) f_2(t) dt \leqslant \sqrt{\int_a^b f_1^2(t) dt \int_a^b f_2^2(t) dt}.$$

We apply it to estimate the quantity $\mathcal{E}$. We express the quantity Δx^2 as

$$\Delta x^2 = 2 \int_0^t \Delta x \frac{d\ \Delta x}{dt} dt \leqslant \sqrt{\int_0^t \Delta x^2 dt \int_0^t \left(\frac{d\ \Delta x}{dt}\right)^2 dt}.$$

Multiplication and division of the right side by the quantity τ yields

$$\Delta x^2 \leqslant \frac{2}{\tau} \sqrt{\int_0^t \Delta x^2 dt \int_0^t \tau^2 \left(\frac{d\ \Delta x}{dt}\right)^2 dt}.$$

Since the integrand functions are positive the inequality is still further strenghtened if the upper limit of integration is increased to infinity:

$$\Delta x^2 \leqslant \frac{2}{\tau} \sqrt{\int_0^\infty \Delta x^2 dt \int_0^\infty \tau^2 \left(\frac{d\ \Delta x}{dt}\right)^2 dt.}$$

From the obvious inequality

$$2 \sqrt{ab} \leqslant a + b$$

- 148 -

it follows that

$$\Delta x^2 \leqslant \frac{1}{\tau} \int_0^\infty \left[\Delta x^2 + \tau^2 \left(\frac{d \Delta x}{dt}\right)^2 \right] dt = \frac{\varepsilon}{\tau}.$$

Hence the deviation of the actual response from the optimum response does not exceed the quantity $\sqrt{\frac{\varepsilon}{\tau}}$:

$$|\Delta x| \leqslant \sqrt{\frac{\varepsilon}{\tau}}$$

Example 3. In a preceding numerical example we found

$$I_{min\ min} \quad \sqrt{2} \approx 1.41 \text{ and } I_{min} \approx 1.53.$$

Then

$$|\Delta x| \leqslant \sqrt{\frac{1.53 - 1.41}{1.41}} = \sqrt{\frac{0.12}{1.41}} \approx 0.292.$$

In other words the selection of system parameters, by means of minimizing the integral performance index,

$$I = \int_0^\infty (x^2 + \tau^2 \dot{x}^2) \, dt,$$

guaranties that the transient response of the system does not leave the boundaries of an area between the curves (Fig. 170)

$$x = x(o) \varepsilon^{\frac{-t}{\tau}} + \sqrt{\frac{\varepsilon}{\tau}},$$
$$x = x(o) \varepsilon^{\frac{-t}{\tau}} - \sqrt{\frac{\varepsilon}{\tau}}.$$

If not only the exponential $x = x(o) \varepsilon^{\frac{-t}{\tau}}$ but also any other curve contained inside this area satisfies the technical specifications for the transient response, then the selection of system parameters is completed.

The greater the value of τ, the smaller is the deviation of the actual response from the exponential one, to which the attempt was made to bring the actual response by selecting the parameters. Thus the exponential curve towards which the response is optimized cannot be specified with too small τ.
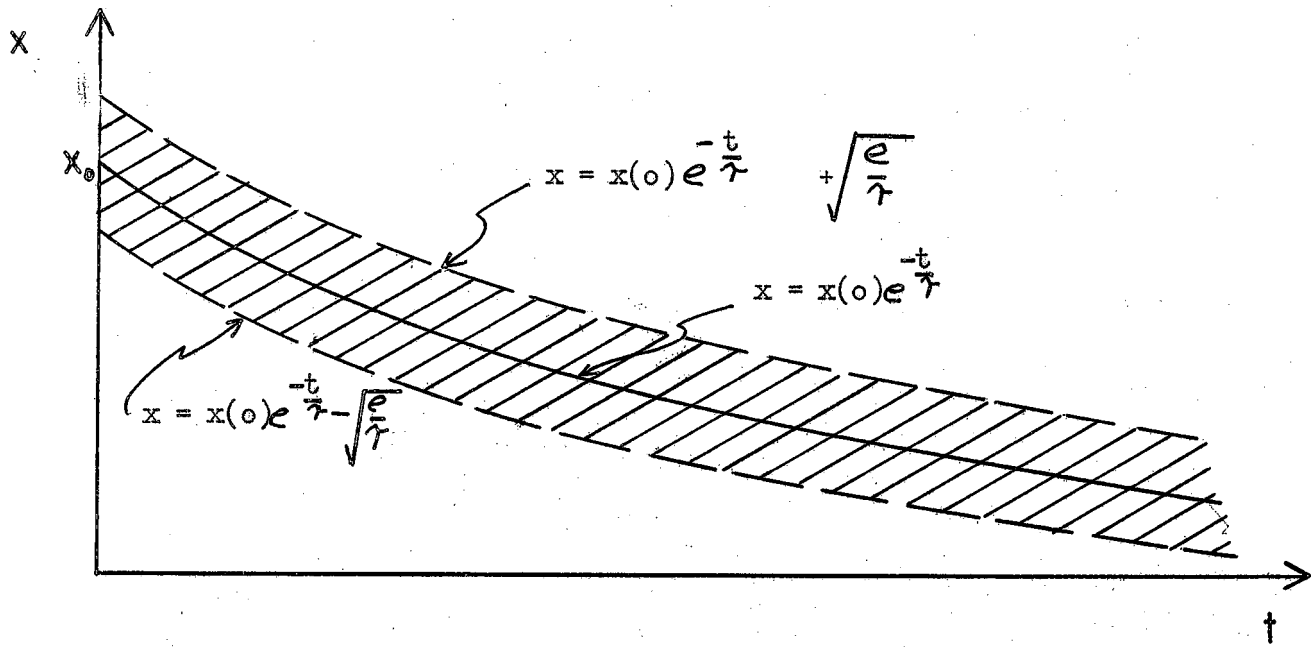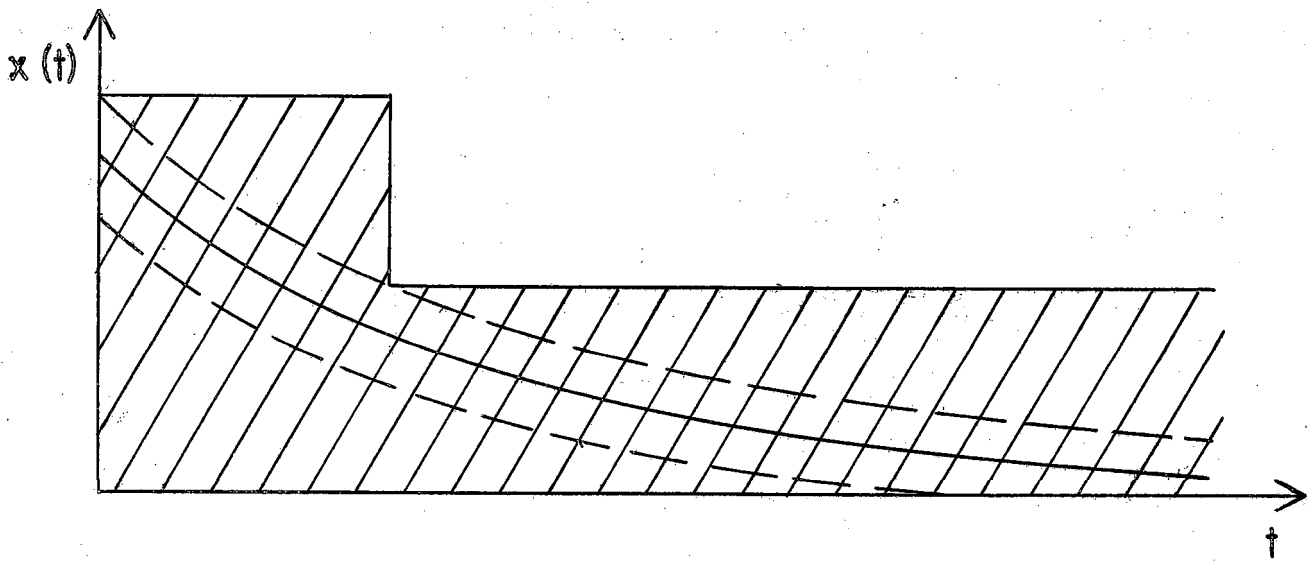
Figure 170

Figure 171

Let the specifications be such that the transient response does not leave the boundaries of the shaded region of Fig. 171.

It is conceivable that the original exponential

$$x = x(o)\epsilon^{-\frac{t}{\tau}}$$

may be selected in such a way that it remains in this region, however, one of the curves

$$x = x(o)\epsilon^{-\frac{t}{\tau}} + \sqrt{\frac{\epsilon}{\tau}} \quad \text{or} \quad x = x(o)\epsilon^{-\frac{t}{\tau}} - \sqrt{\frac{\epsilon}{\tau}}$$

crosses the boundary. In such a case the value τ has to be changed (by selecting a new exponential contained within the shaded region) and the procedure repeated.

It is, however, more convenient not to fix the value of τ at the beginning and to define all unknown quantities as functions of τ. We will illustrate this procedure by means of an example.

Example 4. Referring again to Example 2 we will not define h and τ up to the end but rather we will find them by satisfying the specifications in the best manner. For this purpose we will express h_{min}, I_{min}, and $|\Delta x|$ as function of τ.

We return to the differential equation

$$\frac{d^2x}{dt^2} + h\frac{dx}{dt} + 3x = 0$$

and the original initial conditions

$$x(o) = 1, \quad \dot{x}(o) = 0.$$

Let us select the integral performance index

$$I = \int_0^\infty V\,dt$$

where

$$V = x^2 + \tau^2 \dot{x}^2.$$

In this case

$$A_1 = 1, \quad A_2 = \tau^2, \quad x_1(o) = 1, \quad x_2(o) = 0.$$

Substituting these values into the previously determined equations we get:

$$B_{x_1 x_2} = \frac{1}{3}, \quad B_{x_2} = \frac{\tau^2}{2h} + \frac{1}{6h}, \quad B_{x_1} = \frac{h}{6} + \frac{3\tau^2 + 1}{2h}$$

and consequently

$$I = B_{x_1} x_1^2(o) = \frac{h}{6} + \frac{3\tau^2 + 1}{2h}.$$

Taking the derivative $\frac{dI}{dh}$ and setting it equal to zero we find h_{min} as a function of τ^2:

$$h_{min} = \sqrt{9\tau^2 + 3}$$

Substituting h_{min} into the above determined expression for I we get:

$$I_{min} = \sqrt{\tau^2 + \frac{1}{3}}$$

However, $I_{min\,min} = \tau\, x_1^2(o) = \tau$, since $x_1(o) = 1$. Hence

$$|\Delta x| \sqrt{\frac{\epsilon}{\tau}} = \sqrt{\sqrt{1 + \frac{1}{3\tau^2}} - 1}.$$

Now by varying τ, we find such $\tau = \tau*$ that both curves

$$x = x(o)\epsilon^{-\frac{t}{\tau}} + \sqrt{\sqrt{1 + \frac{1}{3\tau^2}} - 1}$$

and

$$x = x(o)\epsilon^{-\frac{t}{\tau}} - \sqrt{\sqrt{1 + \frac{1}{3\tau^2}} - 1}$$

do not leave the boundaries of the shaded area of Fig. 171.

Then

$$H = \sqrt{9\tau*^2 + 3}$$

is the desired value of h.

e) Computional Procedure

The selection of system parameters, following from the integral per-

formance index

$$I = \int_0^{\infty} (x^2 + \tau^2 \dot{x}^2) \, dt,$$

can be accomplished by two computational procedures.

The first procedure:

1. The constant τ is selected in such a way that the exponential

   $$x^* = x(o)\epsilon^{-\frac{t}{\tau}}$$

   would meet the specifications under the given initial conditions $x(o)$ (which depend upon the initial disturbances).

2. $U(o)$ is determined as a function of the variable system parameters (i.e., parameters which have to be chosen).

3. $U(o)$ is minimized; i.e., the variable parameters are adjusted to yield the absolute minimum of this function, $U_{min}(o)$.

4. $\mathcal{E} = U_{min}(o) - \tau x^2(o)$ is determined and $\sqrt{\frac{\mathcal{E}}{\tau}}$ calculated.

5. The region bounded by the curves $x = x^* + \sqrt{\frac{\mathcal{E}}{\tau}}$ and $x = x^* - \sqrt{\frac{\mathcal{E}}{\tau}}$ is constructed.

The system parameters are determined when every curve inside this region meets the specifications. If this is not the case, the procedure is repeated with a smaller τ. In cases where it is not possible to choose τ such that every curve in the indicated region meets the specifications, it is necessary to change other fixed system parameters and again repeat the procedure. The second procedure:

1. Considering τ as one of the variable parameters, $U(o)$ is determined as function of τ and other variable parameters.

2. This function is minimized with respect to the variable parameters, i.e., the values of the variable parameters which yield

a minimum value of U(o) are determined while U(o) is a function of τ .

3. The constant ε is determined, as a function of τ .

4. The regions bounded by the curves

$$x = x(o)_\varepsilon {}^{\frac{-t}{\tau}} + \sqrt{\frac{\varepsilon}{\tau}}$$

and

$$x = x(o)_\varepsilon {}^{\frac{-t}{\tau}} - \sqrt{\frac{\varepsilon}{\tau}}$$

are constructed in the x, t plane for different values of τ .

If it is possible to find a $\tau = \tau *$ such that any curve in this region meets the specifications of the transient response, the values of variable parameters are considered to be optimal at $\tau = \tau *$.

The selection of parameters based on the integral performance indices is considerably more reliable (trustworthy) than the selection based on the degree of stability consideration. The computations required with the utilization of integral performance indices are, however, more cumbersome.

Nevertheless, except in cases where the system performance specifications are expressed in terms of the response equations (transfer function), the integral performance indices frequently represent the simplest way of selecting the optimal parameters.