7959 R/u

C-L-Chen

[redacted]

Coordinated Science Laboratory

UNIVERSITY OF ILLINOIS – URBANA, ILLINOIS

A METHOD OF ADAPTIVE CONTROL FOR HIGH-ORDER SYSTEMS

Prepared by
E.A.Huber

REPORT R-121 AUGUST 1960

The research reported in this document was made possible by support extended to the University of Illinois Coordinated Science Laboratory jointly by the Department of the Army (Signal Corps), Department of the Navy (Office of Naval Research), and the Department of the Air Force (Office of Scientific Research, Air Research and Development Command) under Signal Corps Contract DA-36-039-SC-85122.

ACKNOWLEDGMENT

The author is grateful to Dr. M.E. Van Valkenburg for his advice and encouragement throughout the preparation of the manuscript, and also to Dr. J.B. Cruz for his advice and helpful criticism in many discussions.

Appreciation is also extended to members of the Adaptive Systems Group for help in computer programming and many interesting discussions.

ABSTRACT

A method of adaptive control is developed which is invariant regardless of the complexity of the system under control. The adaptive control system using this method is a modification of the M.I.T. model-reference system. It is assumed that the type of input signal is known and that it is a pulse or step function. The desired state of the system under control is specified by the appropriate locations of its poles and zeros. Two identical cut-off networks are used whose impulse responses have minimum variance about the time of the maximum. One of these pulse-like outputs is used as a reference for error measurements. The other output is shown to be dispersed when the transfer function of the control system does not meet the specifications. The delay time and width of the response are used as measures of the dispersion, and the necessary adjustments are found from the evaluation of these two error measurements.

The method is applied to a practical pitch-rate control system which is essentially a fourth-order system. In an experimental test of the theory, twelve applications of the input signal were required to set all four poles within a small specified area of the s plane from extreme initial positions.

CONTENTS

ILLUSTRATIONS

ILLUSTRATIONS (continued

1. INTRODUCTION

In the conventional control system, the sensitivity of the system transmission to changes in the parameters of the controlled process or vehicle is reduced by introducing feedback and operating the system with high open-loop gain. The price paid for the reduction in sensitivity is the possible creation of a stability problem. Moreover, large parameter changes require a large gain-bandwidth product for the compensating networks in order to keep the transmission insensitive to these changes. An example of a vehicle having large parameter changes is the present-day aircraft or missile which experiences significant changes in its characteristics or transfer function because it is operating through great extremes of flight conditions. The parameter changes may be so great that excessive gain-bandwidth products would be required for the conventional feedback control system; therefore, changes in the compensating networks are required to maintain the desired aircraft response throughout the entire flight regime[17].

These parameter changes may be made in a predetermined fashion based upon instantaneous air-data measurements, or they may be made automatically by a self-adaptive control system. In order to pre-program the desired changes, very accurate and detailed information about the aircraft characteristics is required for the entire flight regime. The self-adaptive control system eliminates the need for continuous air-data measurements and allows for some ignorance about the aircraft characteristics so that much less flight testing is required. As Horowitz[22] has pointed out, however, one should always weigh the cost of the equipment needed for the self-adaptive features against the cost of the huge gain-bandwidth product required for the conventional feedback control system.

1.1 Definition of a Self-Adaptive Control System

Following Gregory[17], we define a self-adaptive control system as one which has the capability of changing its parameters through an internal process of measurement, evaluation, and adjustment to adapt itself to a changing environment, either external or internal to the process or vehicle under control. It is assumed that the changing environment is manifested by a change in the system parameters.

The requirements for a self-adaptive control system will be explained by use of the generalized model in Figure 1. As mentioned above, three internal operations (measurement, evaluation, and adjustment) are required in a self-adaptive control system. It is assumed that the transfer function of the process or vehicle will vary with time, so the measurement phase is required to identify the present state of the system. This measurement should be performed without disturbing the normal operation of the system. Thus, the common laboratory procedure of injecting pulses or sign waves at the input is not desirable. The types of measurements needed depend on the evaluation scheme used.

The second step is to evaluate the results of the measurement. One must decide whether or not the present state of the system will give the desired response. The output of the evaluation equipment, therefore, will be a figure of merit that will contain enough information to decide whether or not a parameter adjustment is needed and, if it is needed, the direction and magnitude of the necessary adjustment. The parameter adjustment equipment uses the information from the evaluation equipment to make the proper adjustment. In some of the existing systems two of these operations are combined.

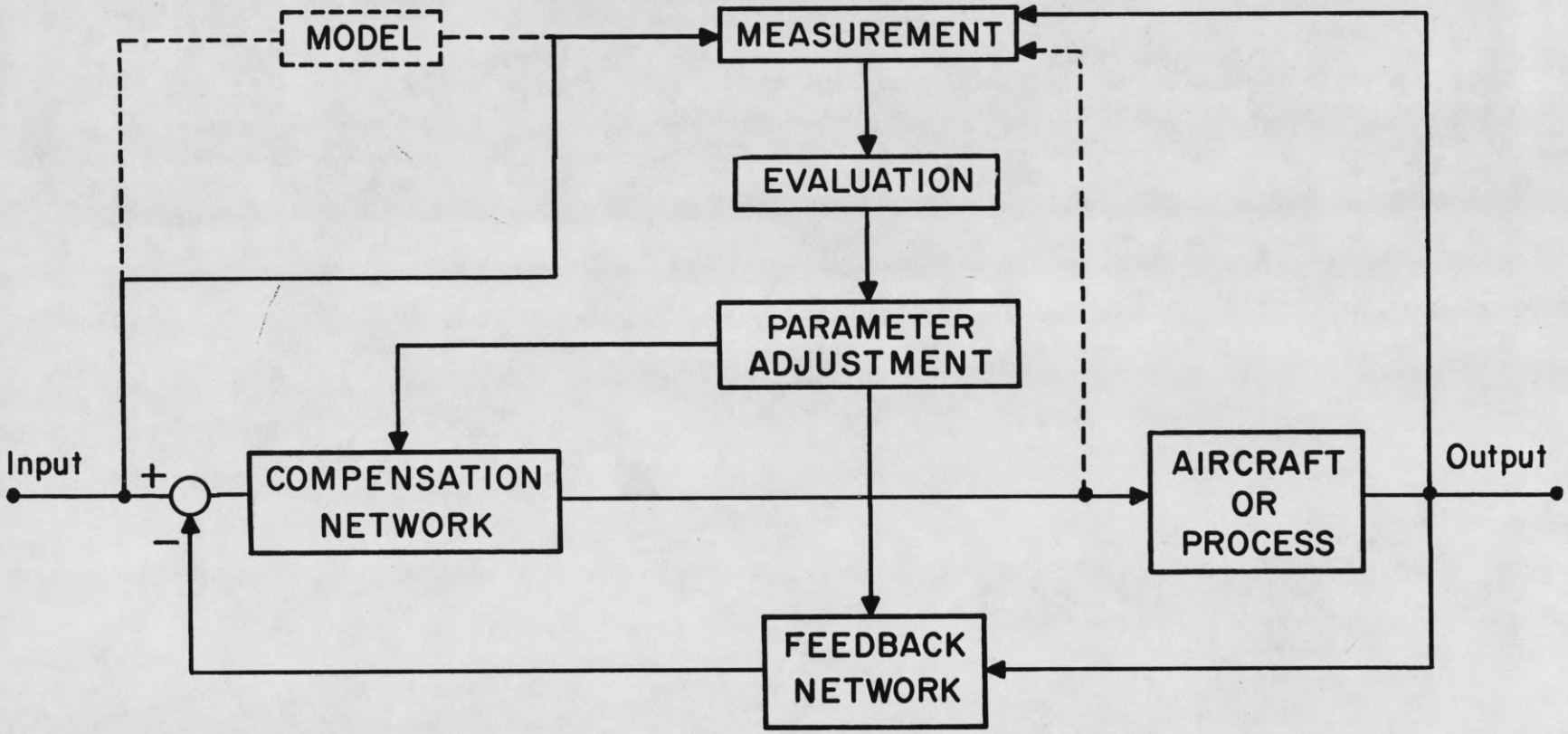


Figure 1. General block diagram for a self-adaptive control system

1.2 Summary of Existing Systems

In the M.I.T. model-reference system[17,29,39], the output of the closed-loop system is compared with the output of the model of the desired system, and an error signal is formed. The error criteria involve nulling the integral of the error and the integral of the absolute value of the error with the latter criterion exerting overriding control. The model-reference system is not restricted to a second-order system; however, the number of error criteria must, in general, be as large as the number of adjustable parameters. The selection of suitable error criteria requires considerable ingenuity. Since all of the required information is obtained from the normal input signal, no test signal is needed.

The General Electric system[17,29] uses a frequency servo to adjust the gain of the compensation network until a preset value of the undamped natural frequency, ω_n, for the control poles is achieved. The system is designed for a particular input signal and may become unstable for a different input signal[29]. Although the procedure is restricted to second-order systems, it is simple to implement and quickly converges to the desired parameter values.

In the Sperry system[17,29,31] a low-level test pulse is applied to the system, and a figure of merit is computed from the number of phase reversals in the signal out of the compensation network. This figure of merit is an indication of the damping ratio, ζ, of the control poles. Again, the procedure is restricted to second order systems, but the instrumentation is simple to implement.

The Aeronutronic system[17,29] obtains the impulse response by adding low-level white noise to the normal signal and using a correlation scheme. This scheme requires complex instrumentation. This part of the method will handle

any order system provided the number of calculated points is increased with the order of the system. A figure of merit for the damping ratio, ζ, is calculated from the impulse response. The figure of merit calculation is restricted to second-order systems.

1.3 Statement of the Problem

Many high-order control systems can be approximated as second-order systems if the same pair of complex poles are predominant for all practical values of the variable parameters[6]. On the other hand, there are high-order systems for which the second-order approximation is not valid. If there is more than one pair of poles close to the $j\omega$ axis or if the parameter variation causes different pairs to predominate at different times, the second-order approximation is not adequate. The problem treated in this paper is that of the high-order system for which the second-order approximation is not valid.

As indicated by the brief survey of existing systems, a simple time-domain measurement is often sufficient to give the location of the dominant poles in the complex frequency plane for a second-order system. The more exact time measure of the impulse response gives enough information to identify any order system, but this information is difficult to use. The servo engineer would generally prefer to know the positions of the poles and zeros of the transfer function in the complex frequency domain. Thus, approximations are made from the time-domain measurements to give complex frequency-domain information. Although these approximations are simple and adequate for the second-order system, they become complicated, inaccurate, or even invalid for the higher-order systems. The transfer function can be obtained by taking the Fourier transform of the impulse response, but this operation requires an

integration with an infinite limit. For high-order systems, then, it seems advantageous to work in the complex frequency domain.

In the M.I.T. model-reference system, the specifications are made in the complex frequency plane by specifying a transfer function which is a model of the desired system. The error measurements are made in the time domain, but the error criteria relate directly to the adjustments needed to obtain a satisfactory pole-zero configuration. This scheme will be extended in this paper to handle the high-order system in an orderly fashion.

The problem treated in this paper, then, is the development of a method of adaptive control which will adjust a high-order control system to meet a given pole-zero specification without using the second-order approximation. The method will be incorporated into a system called the modified model-reference system.

Some of the disadvantages of the previous systems will be alleviated, but some of them will also apply to the modified model-reference system. The system will not require a test signal, but it will be sensitive to the type of input signal encountered. The complexity of the method and the time required for convergence will increase with the order of the closed-loop system.

The following assumptions are made:

1. The initial closed-loop transfer function is approximately known.
2. The desired locations of the closed-loop poles and zeros are specified.
3. The input signal is either a pulse or a step function.
4. Adjustable parameters are available which give somewhat independent control over the poles of interest.

This paper is divided into six major sections. Network concepts needed for the subsequent discussion will be presented in Chapter 2. The modified model-reference system will be explained in detail in Chapter 3. The input signal will be assumed to be an impulse function to simplify the explanation; then, the step function input will be discussed. The method will be applied to a practical aircraft pitch-rate control system in Chapter 4. In Chapter 5 the modified model-reference method will be compared with other possible approaches to the pitch-rate control problem. A summary and discussion of further problems is given in Chapter 6.

2. PERTINENT TOPICS FROM NETWORK THEORY

2.1 Time-Variant Networks

The control systems of interest in this paper are those which need self-adaptive features because some of the aircraft or process parameters are changing with time. The response of the system is related to the input signal through a linear differential equation with variable coefficients. If the input and output variables of a stable system are related by a linear differential equation, Cruz[11] has shown that the system function cannot have singularities that vary with time. In general, the system function of such a system may contain an infinite number of poles. Only a few of the residues of these poles are dominant at a specific time; at a later time a different set may become dominant. Thus, if the parameters are slowly varying with time, it may appear that the number of poles remain fixed and their locations vary with time.

Zadeh[40] has shown that when the coefficients of the fundamental differential equation relating the input and response functions vary slowly with time, the frozen system function may be regarded as a <u>first approximation</u> to the actual system function of a variable network. The differential equation relating the input, x(t), and the output, y(t), of the closed-loop system is written as:

$$\left[a_n(t)\frac{d^n}{dt^n}+\ldots+a_1(t)\frac{d}{dt}+a_o(t)\right]y(t)=\left[b_m(t)\frac{d^m}{dt^m}+\ldots+b_1(t)\frac{d}{dt}+b_o(t)\right]x(t) \qquad (2.1)$$

The frozen system function, $H_F(j\omega, t)$, is defined as:

$$H_F(j\omega, t) = \frac{b_m(t)(j\omega)^m + \ldots + b_1(t)(j\omega) + b_o(t)}{a_n(t)(j\omega)^n + \ldots + a_1(t)(j\omega) + a_o(t)} \tag{2.2}$$

The frozen system function will, in general, have singularities that vary with time.

How slow must the parameters vary in order for the frozen system function to be a good approximation to the actual system function? This question is best answered by using Kailath's[23] development of the system variation. The system function for a time-variant network is defined as:

$$H(j\omega, t) = \int_o^\infty h(\tau, t)\, e^{-j\omega\tau}\, d\tau \tag{2.3}$$

where $h(\tau, t)$ is the response measured at time t due to an impulse applied τ seconds earlier. This equation is equivalent to:

$$H(j\omega, t) = \left.\frac{y(t)}{x(t)}\right|_{x(t)=e^{j\omega t}} \tag{2.4}$$

where x(t) and y(t) are the input and output functions, respectively. Taking the Fourier transform of $H(j\omega, t)$, we obtain a bi-frequency response function

$$H(j\omega, j\mu) = \int_{-\infty}^{\infty} H(j\omega, t)\, e^{-j\mu t}\, dt = \int_{-\infty}^{\infty}\int_o^\infty h(\tau, t)\, e^{-j\omega\tau}\, e^{-j\mu t}\, d\tau dt \tag{2.5}$$

The symbol μ is the variable corresponding to the system variation and ω is the variable corresponding to the output variation.

Kailath gives a qualitative concept of μ, the frequency-domain measure of the variation of the system. He points out that if μ were confined to

low values the system would be slowly varying; it would vary rapidly if there were high frequencies present in the μ domain. Let us make the concept quantitative by stating that we will consider the system to be slowly varying if the bandwidth in the ω domain for all values of t is ten times the bandwidth in the μ domain. In other words, the impulse response of the system dies out before there is a significant change in the system's parameters. Under these conditions, it will be considered valid to consider poles varying with time.

2.2 Pulse Dispersion

In the study of the propagation of electromagnetic waves through a medium, the concepts of the various velocities of propagation have been given a thorough treatment. Brillouin[5] has shown that a precursor always arrives at the wave-front velocity which equals the velocity of light. For fields which are periodic in space, the surfaces of constant phase are propagated at the phase velocity

$$v_{\varphi} = \frac{\omega}{k} \tag{2.6}$$

where k is the wave number and ω is the radian frequency. The simplest treatment of group velocity is based on the consideration of two steady-state sinusoidal waves of equal amplitude and slightly different frequencies. The "beats" or "groups" formed by the sum of these two waves are propagated at the group velocity

$$v_g = \frac{d\omega}{dk} \tag{2.7}$$

A "wave packet" composed of waves within an infinitely narrow region of the spectrum, Δk, is also propagated at the group velocity. As the interval, Δk, is increased, the spread in phase velocity of the harmonic components

in a dispersive medium becomes more marked; the packet is deformed rapidly, and the group velocity as a velocity of the whole loses its physical significance[32]. If the dispersion is normal, the group velocity is less than the phase velocity; if the dispersion is anomalous, the group velocity is greater than the phase velocity. In the neighborhood of an absorption band, the group velocity can become negative near the resonant frequency.

Hartley[21] gives a good analogy between the velocities of propagation in a medium and the time delays in a linear system. He points out that the group velocity corresponds to group or envelope delay.

$$\tau_g = \frac{-\, d\ \varphi(j\omega)}{d(j\omega)} \tag{2.8}$$

The symbol $\varphi(j\omega)$ is the phase of the transfer function of the system and ω is the radian frequency. The analogue of a medium having normal dispersion is a "damped" system which exhibits a gradual variation of both phase and attenuation with frequency. A "resonant" system exhibits characteristics analogous to anomalous dispersion. He also points out that if the attenuation is uniform over a band and infinite outside the band, the group delay is a valid measure of the arrival time of the envelope maximum. When the attenuation varies rapidly with frequency, the group delay gives the delay in the maximum value of the envelope, but this maximum is so flat that the interpretation of the results is very difficult.

Any distortion in either the phase or attenuation characteristics of a filter which is flat in attenuation and linear in phase will cause dispersion of the impulse response. Di Toro[12] has treated attenuation distortion of the form $(a\omega)^m$ and phase distortion of the form $(b\omega)^n$. He has

shown that in the absence of attenuation distortion the group delay gives the time when the impulse response has a specific frequency for the type of phase distortion noted above. Either infinite order attenuation distortion[18] or infinite order phase distortion[13] of the type noted above results in the (sin x)/x-type output.

Thus, group delay can be used as a measure of the time at which energy will be received at the output of a filter provided there is not a rapid variation in attenuation. If there is a rapid variation in attenuation, there will be additional dispersion, and energy will be received at times not accountable by the group-delay measure alone.

The "ideal" cut-off network[18] has a rectangular magnitude characteristic and linear phase characteristic; i.e.,

$$|H(j\omega)| = \begin{cases} K, & |\omega| < |\omega_c| \\ 0, & |\omega| > |\omega_c| \end{cases} \tag{2.9}$$

$$\varphi(j\omega) = - j\omega b, \quad |\omega| < |\omega_c|$$

and the impulse response is

$$h(t) = \frac{k\omega_c}{\pi} \left[\frac{\sin \omega_c (t-b)}{\omega_c (t-b)} \right] \tag{2.10}$$

This pulse occurs at the group delay time which is the negative of the slope of the total phase function

$$\tau_g = - \frac{d\, \varphi(j\omega)}{d(j\omega)} = b \tag{2.11}$$

If a cut-off network has a magnitude characteristic which is approximately rectangular and a phase characteristic which is linear over the pass band, its impulse response will be a slightly wider pulse than that mentioned above, but it can still be used as a reference signal. A cut-off network whose impulse response makes a good reference signal has the magnitude characteristic shown in Figure 2a, and the phase characteristic shown in Figure 3a. Its impulse response is shown in Figure 4a.

Either amplitude or phase distortion will cause dispersion of the pulse. The distortion caused by adding poles or zeros in the pass band of the network will cause a type of dispersion that can be predicted. In particular, when a pair of complex conjugate zeros are added to the network at the complex frequency, $s_1 = -\sigma_1 \pm j\omega_1$, the magnitude function becomes like that shown in Figure 2b, and the phase function becomes like that shown in Figure 3b. The average slope of the phase function is less than that of the reference because the two zeros cancel π radians of the phase at infinite frequency. More significant, however, is the reduction in the slope of the phase function near ω_1, the frequency of the zeros. Since there is little change in the magnitude function in this region, some of the energy will be received at a much earlier time than that of the reference. The rapid change in the magnitude function at higher frequencies causes dispersion in addition to that caused by the phase distortion. The resultant impulse response is shown in Figure 4b.

When a pair of complex poles are added at s_1, the slope of the phase function is increased and the pulse is delayed in time. The magnitude function is shown in Figure 2c, the phase function in Figure 3c, and the

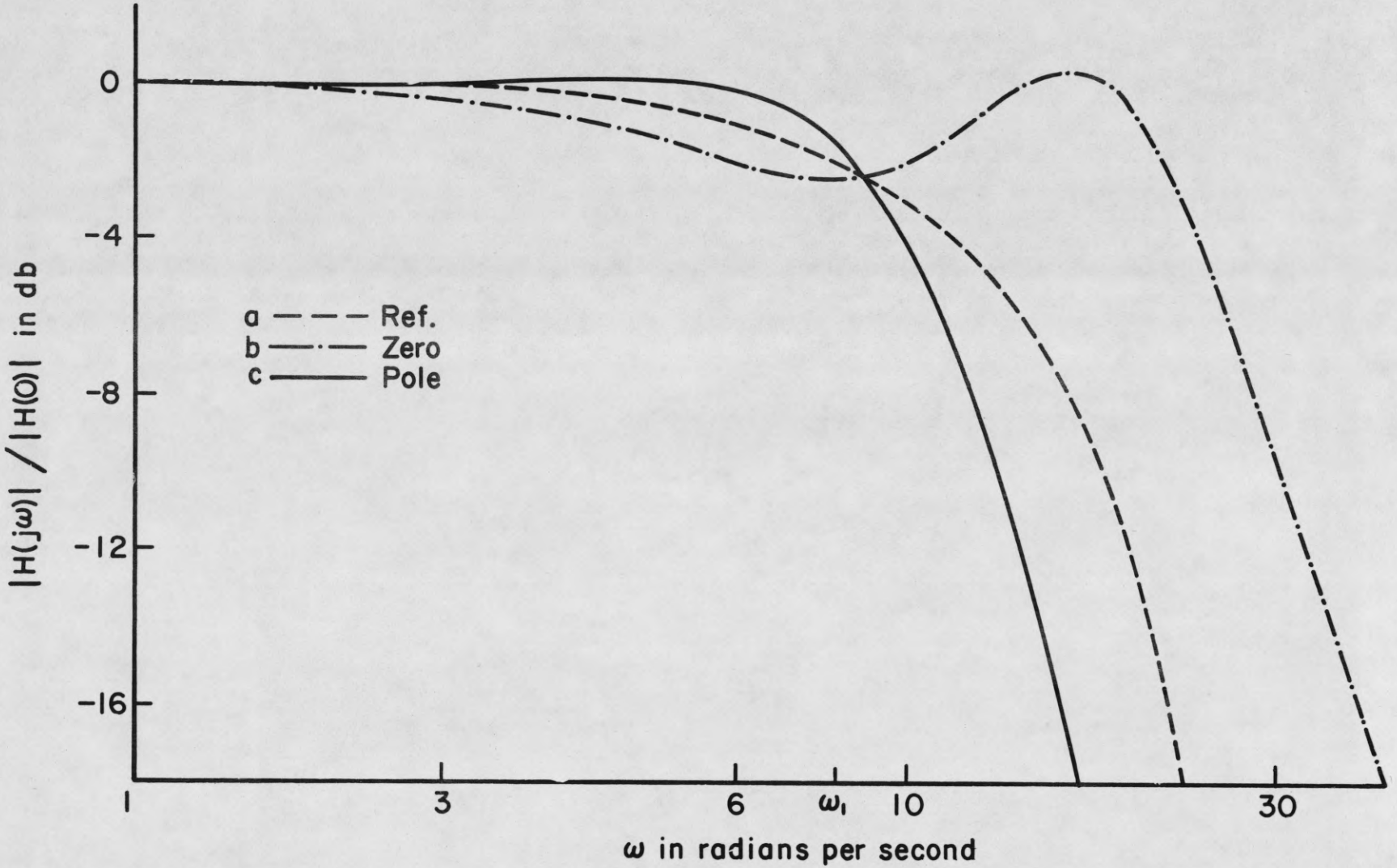


Figure 2. Magnitude characteristics

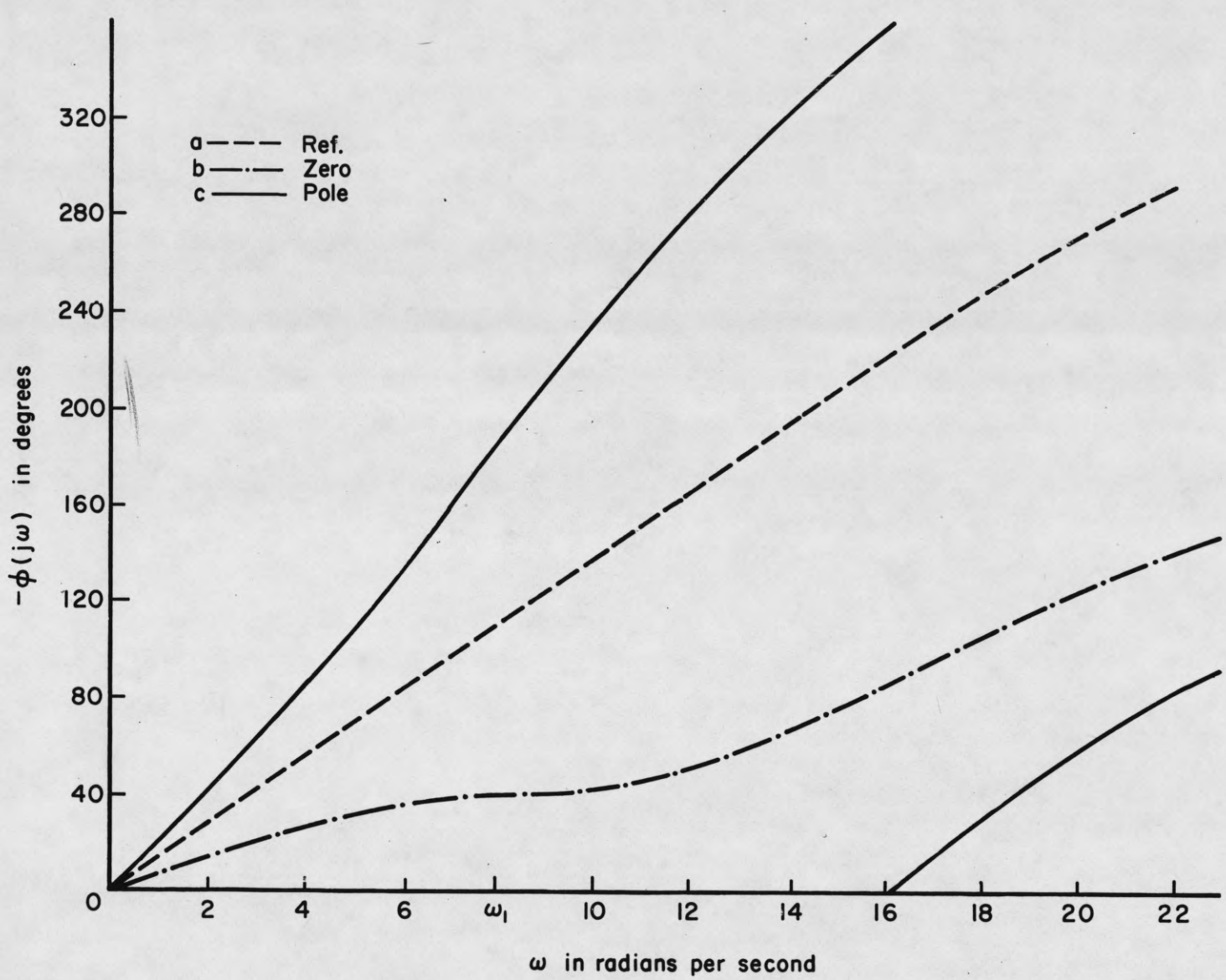
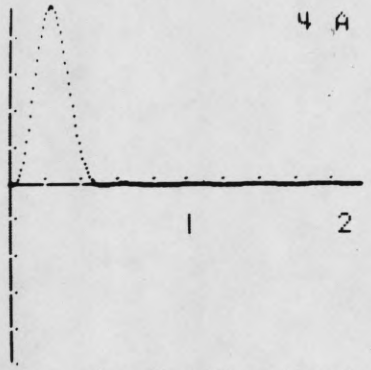
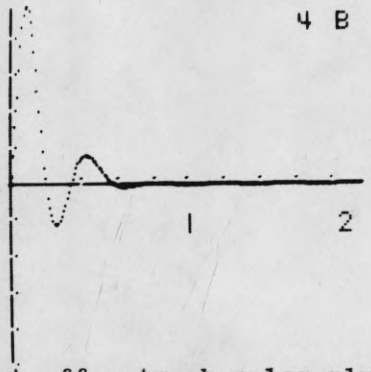


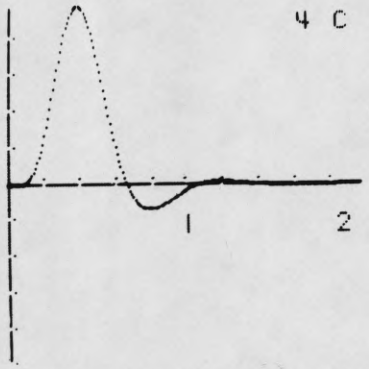
Figure 3. Phase characteristics



a) Poles of cut-off network only



b) Cut-off network poles plus complex zeros



c) Cut-off network poles plus complex poles

Figure 4. Normalized impulse responses

impulse response in Figure 4c. If a pair of dipoles (one pole and one zero close together) are added to the network in the pass region, the dispersion can be used as a measure of error. The error will be minimum when the pole and zero cancel each other. When the zeros are dominant, the pulse is shifted forward in time; when the poles are dominant, the pulse is delayed in time. The greater the distance between the pole and zero of the dipole, the greater will be the dispersion. These concepts will be used in the sections on error measurement and evaluation in Chapter 3.

2.3 Method of Calculating the Impulse Response

The network transfer function

$$H(s) = K \frac{\pi_i (s + z_i)}{\pi_j (s + p_j)} \tag{2.12}$$

is related to the impulse response, h(t), through the inverse Fourier transform:

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(j\omega)\ e^{j\omega t}\ d\omega \tag{2.13}$$

This operation, however, is difficult to perform on the digital computer because it involves either integrating with infinite limits or using finite limits and having a truncation error. For time repetitive signals, spectral sampling[3] may be used to convert from frequency-domain to time-domain functions with a digital computer. The time functions are first resolved into Fourier components. The effect of the system on the input waveform is then found by modifying the magnitude and phase of each harmonic component by the loss and phase of the system transfer function. The new time functions are then added to give the desired output waveform.

The method used in this paper is to expand the transfer function in a partial fraction expansion

$$H(s) = \sum_i \frac{K_1}{s + p_i} \tag{2.14}$$

and sum the inverse Fourier transforms of the individual terms. A transfer function with multiple poles is approximated by one with closely spaced simple poles. The inverse transform of a term with a negative real pole is

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{K_1}{j\omega + a_1} e^{j\omega t} \, d\omega = K_1 e^{-a_1 t} \tag{2.15}$$

The inverse transform of a pair of terms with complex conjugate poles is

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\frac{K_2}{(j\omega + a_2 + jb_2)} + \frac{K_2^*}{(j\omega + a_2 - jb_2)} \right] e^{j\omega t} d\omega = e^{-a_2 t} [2R_2 \cos b_2 t + 2I_2 \sin b_2 t] \tag{2.16}$$

where $K_2 = R_2 + jI_2$ and $K_2^* = R_2 - jI_2$,

The inverse transforms of the individual terms are stored in the computer and the a_i's and b_i's are input data, so the computer simply evaluates the residues and sums terms of the form of Equations 2.15 and 2.16.

2.4 Sensitivity

Bode[4] has defined the sensitivity of the network transmission in terms of the percentage change of the over-all transmission, T(s), and the percentage change in a single variable element, k. It is common practice to use the reciprocal of Bode's sensitivity function as the definition of the sensitivity of T(s) with respect to k[37].

$$S_k^T = \frac{d(\ell n \, T(s))}{d(\ell n \, k)} = \frac{\frac{dT}{T}}{\frac{dk}{k}} \tag{2.17}$$

This sensitivity function applies for incrementally small changes in k.

The block diagram of a system with a variable parameter k can always be put into the form of Figure 5[35]. The symbol $T_o(s)$ represents the leakage transmission with k = 0. The symbols A(s) and B(s) represent transfer functions, and C(s) is the transmission around the element k. In terms of these quantities, the sensitivity function becomes:

$$S_k^T = \frac{1}{1 - k\,C(s)} \left(1 - \frac{T_o(s)}{T(s)} \right) . \qquad (2.18)$$

When the leakage transmission is small with respect to the transmission through the variable element,

$$S_k^T \approx \frac{1}{1 - k\,C(s)} \qquad (2.19)$$

and the loop gain, k C(s), must be large over the frequency range in which low sensitivity is required. Beyond this frequency range, the high loop gain must be reduced gradually (less than 33 db/decade) in order to avoid stability problems. Thus, sensitivity is improved at the price of an increased gain-bandwidth requirement and a possible stability problem.

The sensitivity can be reduced without high gain-bandwidth requirements for special classes of input functions. The sensitivity of the magnitude of the transmission gain can be made small for sinusoidal inputs, for example, by arranging the system so that the real part of the sensitivity function is small[35]. The sensitivity of the magnitude of the transmission gain is reduced at the expense of poor phase sensitivity.

The sensitivity function defined in Equation 2.17 is not valid for large changes in the elements if the over-all transmission is not a linear

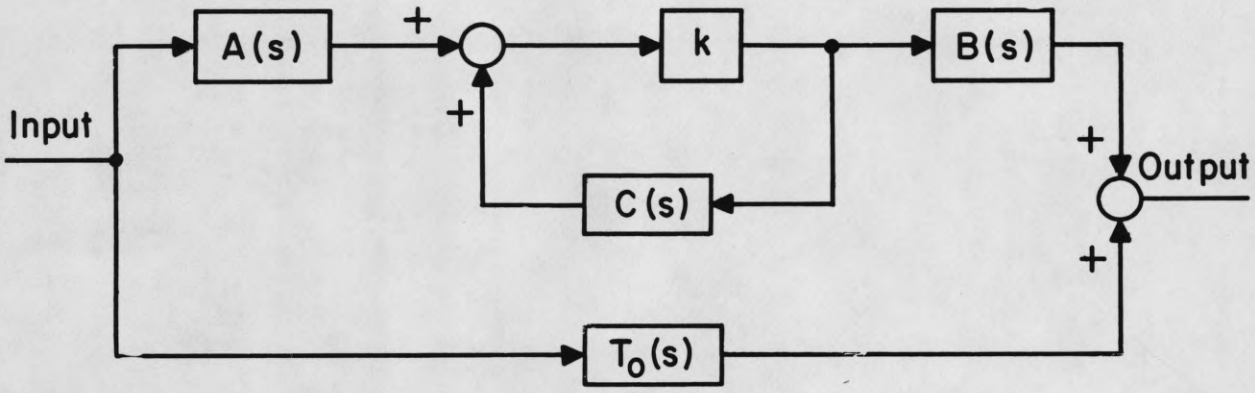


Figure 5. Block diagram of a system with a variable parameter k

function of the variable elements. The best measure of the maximum variation of $T(j\omega)$ for large variations in more than one element is obtained by calculating $T(j\omega)$ over the frequency range of interest for all possible values of the variable elements. Using $T_n(j\omega)$ as the normal transmission gain, we would then determine the maximum deviation from $T_n(j\omega)$ for each frequency as an upper bound on the variation of $T(j\omega)$. When the parameters can vary continuously, it becomes impossible to make the required calculation.

Hakimi and Cruz[20] have found an upper bound for large multiple-parameter variations. This upper bound is easy to calculate, but it may be greater than the least upper bound. They also calculate the variation by picking random values for the parameters within the tolerances allowed. An indication of the variation in $T(j\omega)$ can also be found by letting each variable parameter have only two values, its maximum value and its minimum value. There will be 2^n calculations of $T(j\omega)$ for all the combinations of element values where n is the number of variable parameters. From this set of $T(j\omega)$ calculations, T_m, the one which has maximum deviation from T_n can be found for each frequency. A plot of $\left| 20 \log_{10} |T_m/T_n| \right|$ then gives a measure of the variation of $T(j\omega)$. This quantity will not necessarily be as large as the least upper bound, but it is easy to calculate and useful for the type of assertions made in Chapter 5.

3. MODIFIED MODEL-REFERENCE SYSTEM

As mentioned in the introduction, the modified model-reference system is intended to give complex frequency information about high-order systems in an organized manner. A block diagram of the modified model-reference system is shown in Figure 6. It has two reference networks: a cut-off network and a specification network. The cut-off network limits the response, for measurement purposes, to only that region of the complex frequency plane under investigation and gives a response which makes a good reference for error measurement and evaluation. The specification network acts as the inverse of the model of the desired closed-loop system; i.e., zeros are placed at the desired pole locations and vice versa. The question of practical realizability of the specification network will be discussed later.

When the response of the closed-loop system to poles in the isolated region meets the specifications, the poles and zeros of the closed-loop system in that region will be canceled by those of the specification network, and the inputs to the measurement equipment will be the identical responses of the cut-off networks to the input signal. In essence, the pole locations are set by adjusting the parameters until the poles are canceled by the specification zeros. The method of determining the pole locations of a network by complex-zero inputs has been used by Brussolo[7] and Lendaris and Smith[26]. They obtained arrays of complex zeros by generating two step functions which have adjustable amplitudes and an adjustable delay time between them. The delay time determines the spacing along the imaginary axis, and the amplitude ratio determines the position along the real axis. These complex zeros are varied until the poles of the network are canceled. The

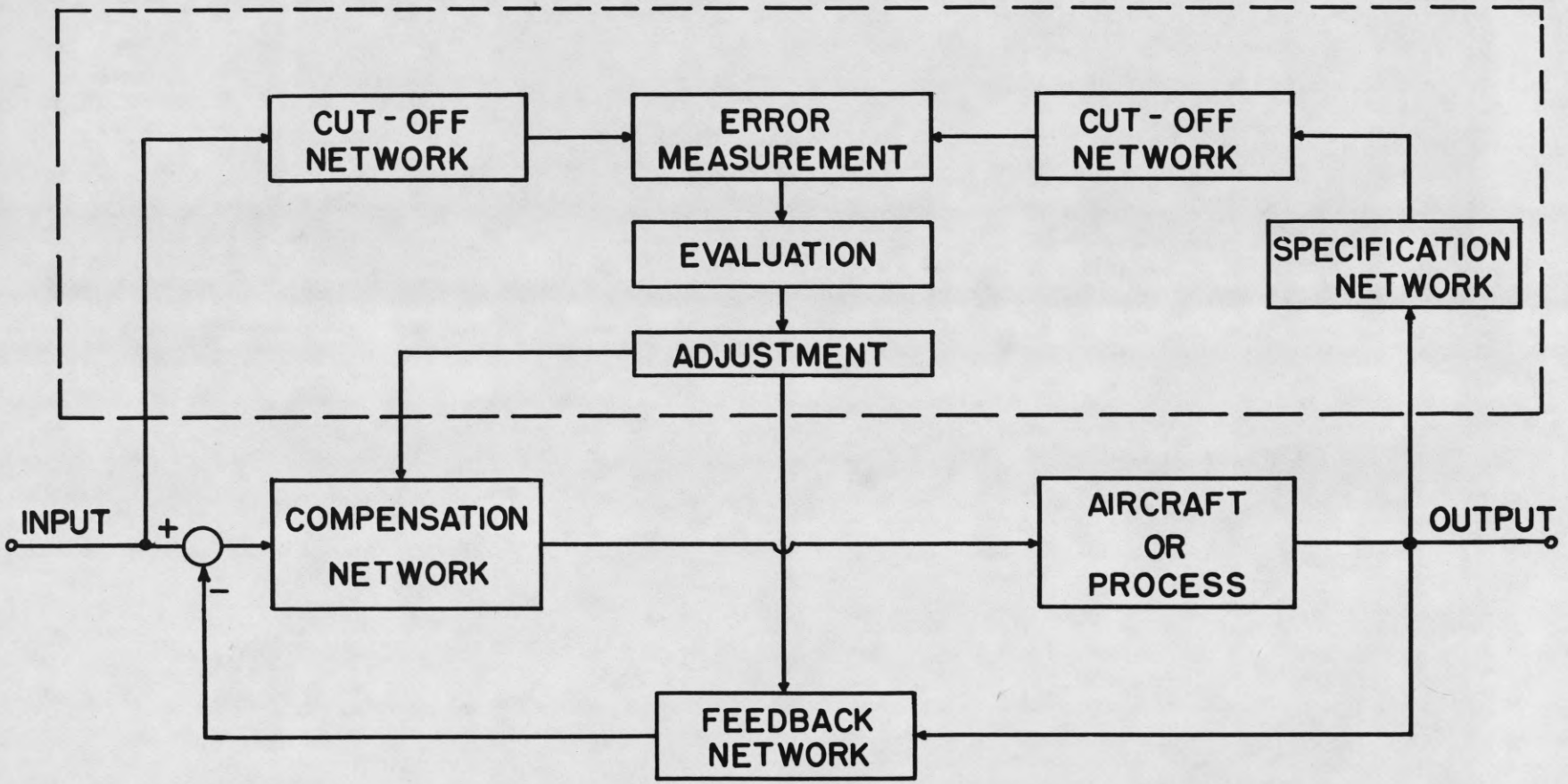


Figure 6. Block diagram of the modified model-reference system

good reference. If the width, Δt, of the pulse, h(t), is taken (after Woodward[38]) to be the width of a rectangle having the same height and area as the $|h(t)|^2$ curve, then

$$\Delta t \triangleq \frac{\int_0^\infty |h(t)|^2 \, dt}{|h(t)|^2_{max}} \tag{3.1}$$

For the band-limited filter, without regard to physical realizability, the minimum Δt is obtained with a rectangular magnitude and linear phase characteristic. The impulse response of such a filter is the (sin x)/x function which has considerable side-lobe structure.

A better measure of pulse width for our purpose is the variance about the time of the maximum, i.e.,

$$\sigma^2 = \frac{\int_0^\infty (t-t_m)^2 \, |h(t)| \, dt}{\int_0^\infty |h(t)| \, dt} \tag{3.2}$$

where t_m is the time when h(t) is maximum. The condition for minimum variance, as will be shown later, also requires a transfer function with a linear phase characteristic.

The magnitude characteristics for the four filters mentioned above for the fifth-order case are shown in Figure 7, and the corresponding phase characteristics are shown in Figure 8. For comparison purposes, the pole locations have been scaled so that the filters all have a half-power point at ten radians per second. The pole locations for these plots are shown in Figure 9, and the corresponding impulse responses are shown

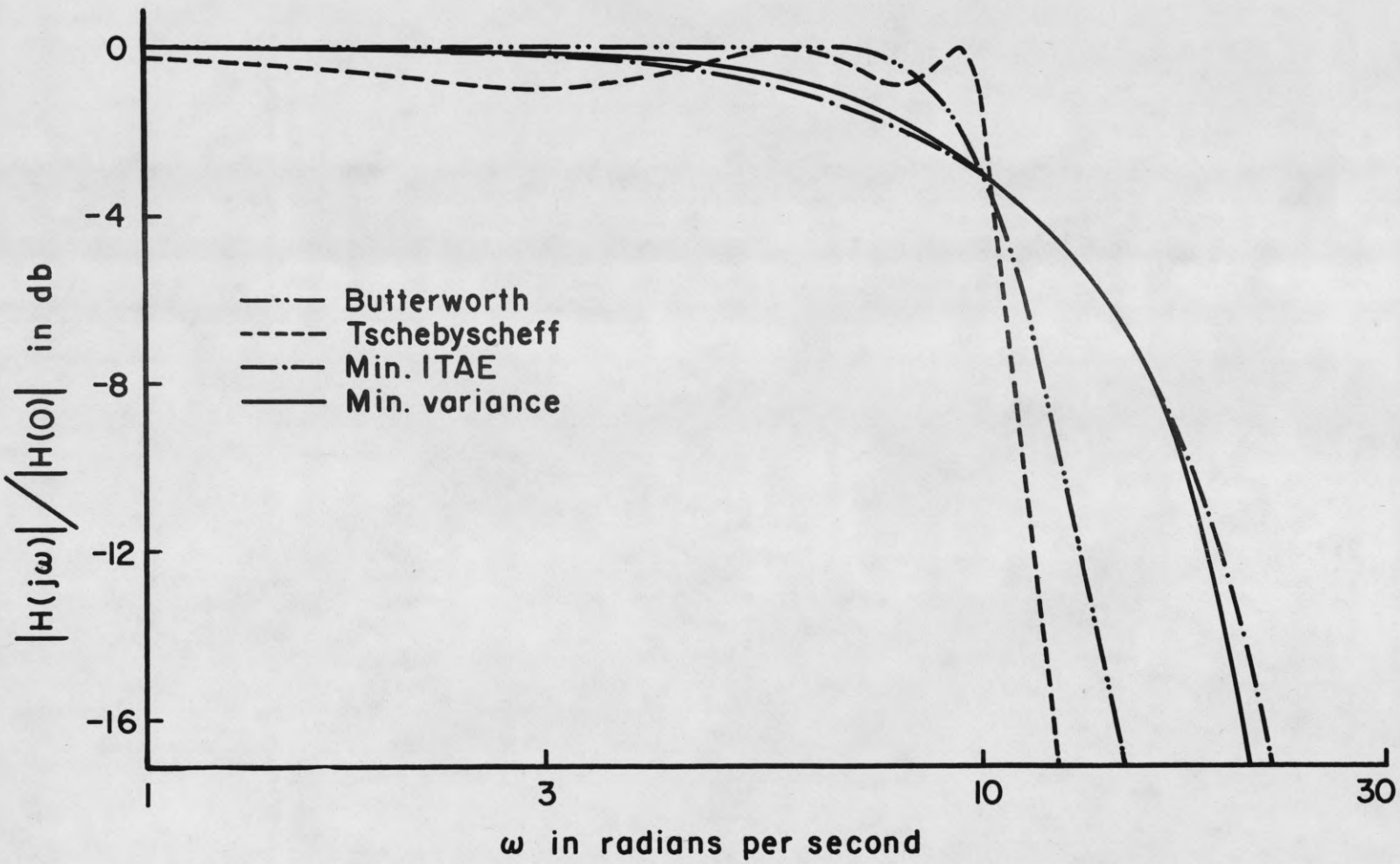


Figure 7. Magnitude characteristics of four fifth-order low-pass filters

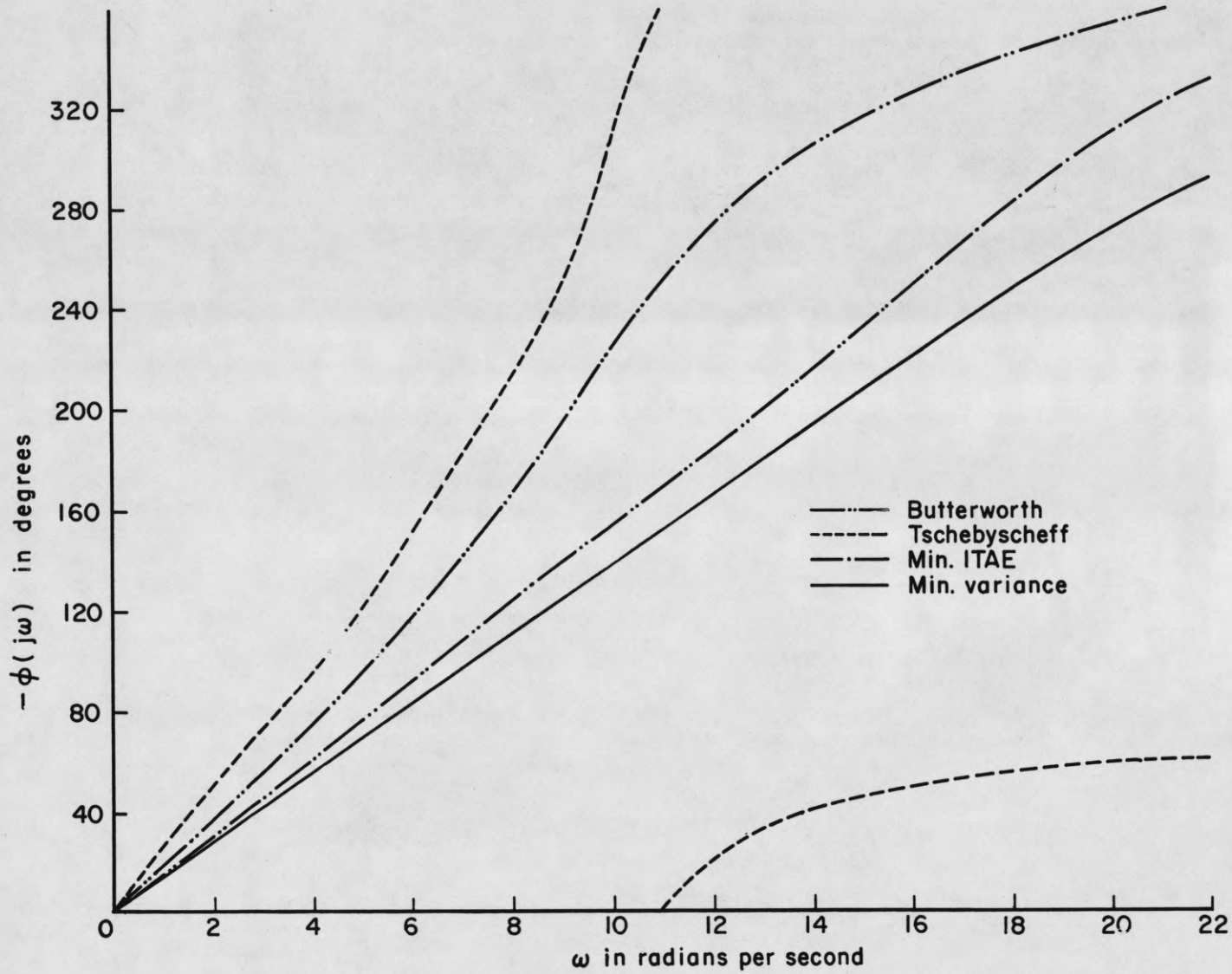


Figure 8. Phase characteristics of four fifth-order low-pass filters

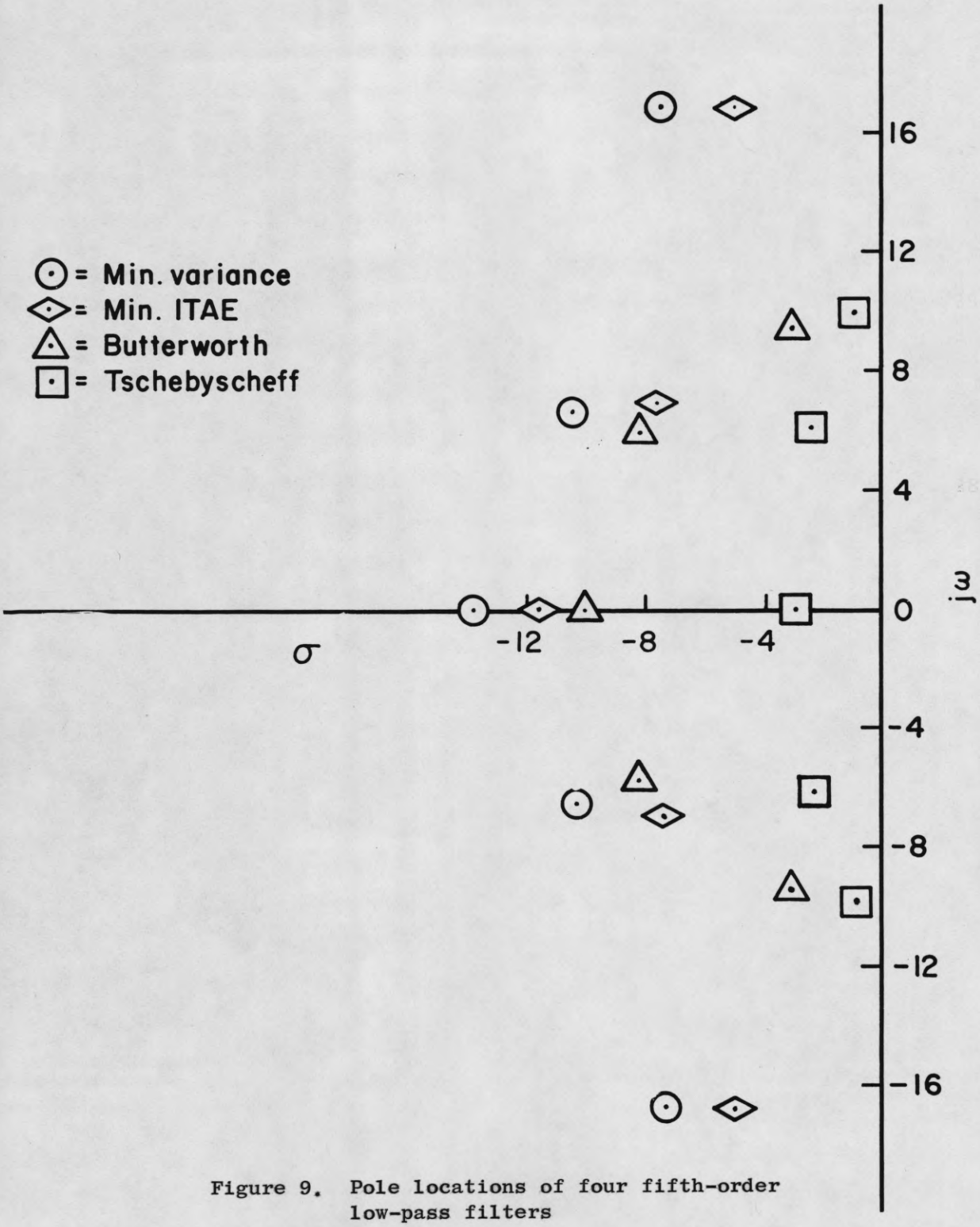


Figure 9. Pole locations of four fifth-order low-pass filters

in Figure 10. The step responses are shown in Figure 11.

The Butterworth filter has a fairly sharp cut-off in the magnitude function, but its phase function is somewhat nonlinear. Although, the Tschebyscheff filter has a very sharp cut-off in the magnitude function, its phase function is very nonlinear. The minimum ITAE filter has poorer cut-off features than the Butterworth filter, but its phase function is linear over the pass band. The minimum variance filter is very similar to the minimum ITAE filter. Since its phase function is linear over a greater frequency band, the slight ringing in the impulse response of the minimum ITAE filter is eliminated. The minimum variance filter will be used as the cut-off network in most of the work that follows; therefore, a description of the method used to obtain this filter is given next.

In order to obtain the minimum variance filter, the number of poles was fixed at five and σ^2, as defined below, was computed on the ILLIAC digital computer.

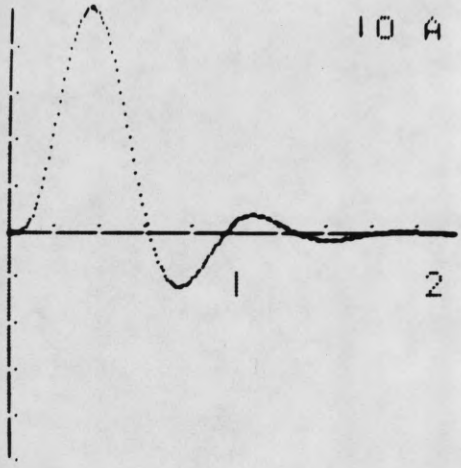
$$\sigma^2 = \frac{\frac{K}{t_m^2} \int_0^{4t_m} (t_m - t)^2 \, |h(t)| dt}{\int_0^{4t_m} |h(t)| dt} \tag{3.3}$$

The symbols are defined as follows:

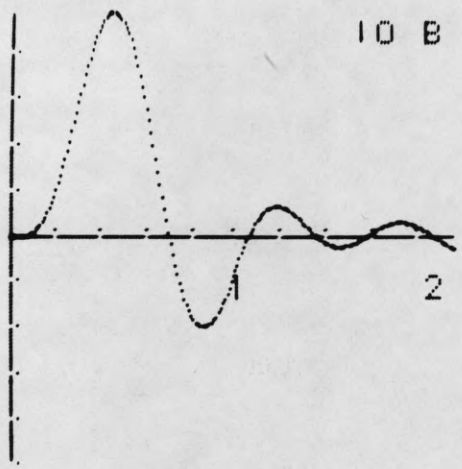
K = constant

t_m = time at which $h(t)$ is maximum

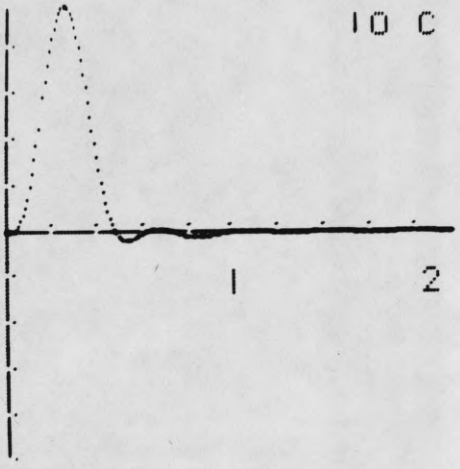
$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{j\omega t} \, d\omega}{[(-\omega^2 + a_1^2 + b_1^2) - j2a_1\omega][(-\omega^2 + a_2^2 + b_2^2) - j2a_2\omega][j\omega - a_3]} \tag{3.4}$$



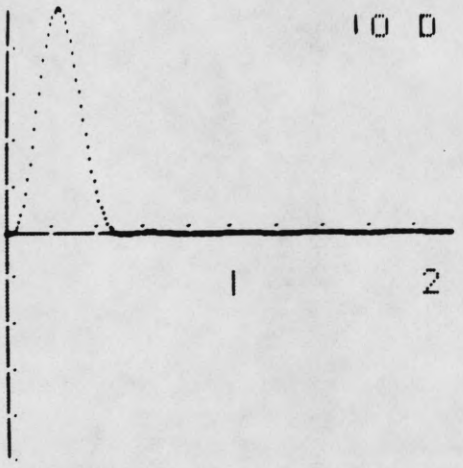
a. Butterworth filter



b. Tschebyscheff filter

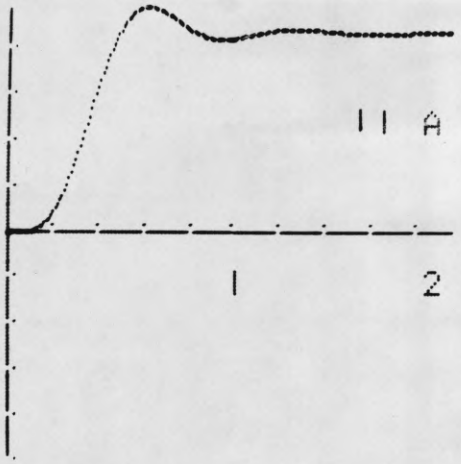


c. Min ITAE filter

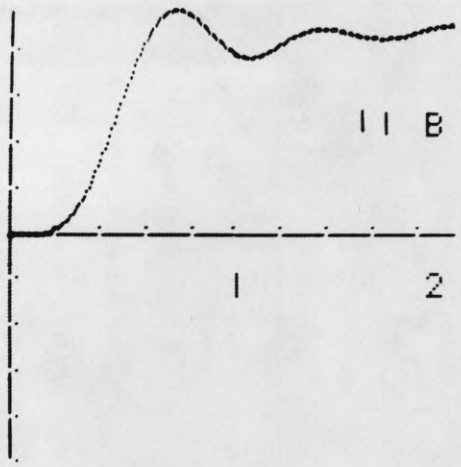


d. Min. variance filter

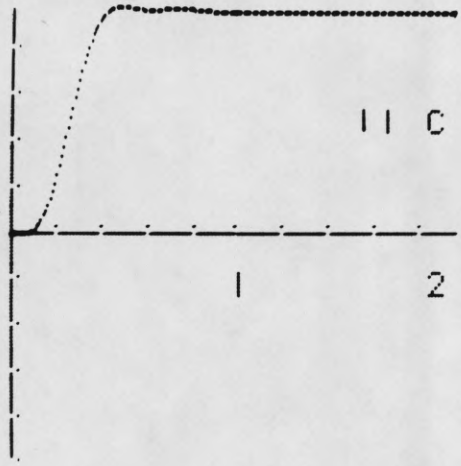
Figure 10. Normalized impulse responses of four fifth-order low-pass filters



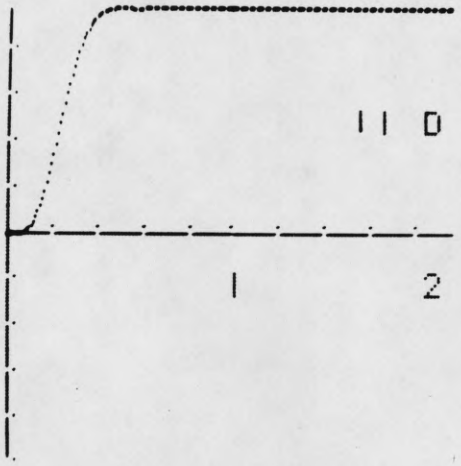
a. Butterworth filter



b. Tschebyscheff filter



c. Min. ITAE filter



d. Min. variance filter

Figure 11. Normalized step responses for four fifth-order low-pass filters

where a_1, a_2, a_3, b_1 and b_2 are the variable parameters that determine the five pole locations. The five parameters were initially set at the values corresponding to the fifth-order Butterworth filter. The parameters were then varied one at a time until σ^2 was minimized for the first mesh size. The mesh size, which determined the size of the variation, was then reduced and a new minimum σ^2 was found. This process was repeated until the pole locations were determined to within three significant figures. The pole locations for a fifth-order transfer function with the half-power point at ten radians per second are:

$$-10.6 \pm j\ 6.81$$

$$-7.61 \pm j\ 17.28$$

$$-14.1 \pm j\ 0$$

The function h(t) was calculated on the computer for 200 points, the last point being $4t_m$. At $4t_m$, $h(t) < 0.01\ h(t_m)$ for all h(t)'s. Integration was performed using Simpson's rule, so the approximation

$$\frac{1}{4t_m} \int_o^{4t_m} f(x)\ dx \approx \frac{1}{3n} (f_o + 4f_1 + 2f_2 \ldots + f_n) \tag{3.5}$$

was used.

$$\begin{aligned} f_k &= (50-k)^2\ \left|h\left(\frac{kt_m}{50}\right)\right| \\ &= \frac{2500}{t_m^2} (t_m - t_k)^2\ |h(t_k)| \end{aligned} \tag{3.6}$$

where

$$t_k = \frac{k\ t_m}{50}$$

A bound on the bandwidth is obtained by normalizing with respect to t_m. If the phase function were perfectly linear over the pass band and the phase function reached - $5\pi/2$ at the cut-off frequency, ω_c, then

$$t_m = - \frac{d\varphi}{d\omega} = \frac{5\pi}{2\omega_c} \tag{3.7}$$

The increment of time between points is fixed at $4t_m/200 = t_m/50$. If the bandwidth increases, t_m decreases, and the points are placed closer together in time; therefore, no reduction in σ^2 is obtained by merely increasing the bandwidth. Otherwise, the variance could be reduced by merely shoving the poles farther out in the complex frequency plane.

The denominator of Equation 3.3 was chosen to normalize the values of $|h(t)|$. As shown in the previous figures, the minimization of the variance resulted in a filter very similar to the minimum ITAE filter. The main difference is that the phase of the minimum variance filter is linear over a larger frequency range.

3.2 Specification Network

The specification network is the inverse of the desired model of the system insofar as it is practically realizable. The modified model-reference method cannot be applied to a non-minimum phase-shift network because it would require the inverse to have poles in the right half plane. For a minimum phase-shift network, the inverse network may require more zeros than poles, so that it will not be practically realizable. This problem can be circumvented by assigning some of the surplus zeros to the all-pole cut-off network in cascade with the specification network.

When the first set of parameters to be adjusted are correctly set,

the poles and zeros of the closed-loop system cancel those of the specification network in the region isolated by the cut-off network. The impulse response of the closed-loop system, specification network, and cut-off network in cascade is essentially that of the cut-off network alone. When the parameters are not correctly set, the poles and zeros of the specification network will not be completely canceled by the zeros and poles of the control system. The impulse response of the closed-loop system, specification network, and cut-off network in cascade is the impulse response of the cut-off network convolved with the inverse transform of a number of dipoles. When this response is compared with the response of the cut-off network alone, there is an error signal which is evaluated to determine in which direction the poles need to be "moved". The parameters are then adjusted until complete cancellation is obtained.

3.3 Error Measurements

Assuming that the system is not operating as specified, the measurement and evaluation of the error increases in complexity with the number of dipoles in the isolated region. The regions isolated by the cut-off networks should be chosen so that only one pair of dipoles (corresponding to one real pole of the system), or two pairs of dipoles (corresponding to a pair of complex poles of the system) are present in the isolated region. After the parameters have been adjusted so that these poles are canceled, the cut-off networks are changed to include a larger region. The case in which there are two complex poles and one real pole in the isolated region can also be handled, but for more complex situations the evaluation phase becomes too complicated and convergence time too long. Some flexibility in meeting these requirements can be obtained by changing the type of cut-off networks

as well as the bandwidth. If the system is such that the poles cannot be isolated one, two, or three at a time by the cut-off networks, then the modified model-reference scheme cannot be applied.

It is assumed, for the general problem, that there are system poles outside the isolated region not canceled by zeros of the specification network, and two error measurements are used: one is used to measure dispersion and the other is used to measure the time delay. The best measure of dispersion is the variance given by Equation 3.2. This measure is somewhat difficult to implement and a cruder measure of dispersion will suffice. The measure actually used is the width of the rectangle with height $|h(t_m)|$ and area $\int_o^{t_c} |h(t)|dt$. The pulse-width error using this measure is denoted $e_{\Delta t}$ and is:

$$e_{\Delta t} = \frac{\int_o^{t_c} |h_l(t)|dt}{|h_l(t)|_{max}} - \frac{\int_o^{t_c} |h_r(t)|dt}{|h_r(t)|_{max}} \qquad (3.8)$$

The symbols in the equation are the following:

$h_r(t)$ is the reference impulse response from the cut-off network alone

$h_l(t)$ is the impulse response of the system, specification network, and cut-off network in cascade

$h(t)_{max}$ is the maximum value of the impulse response

t_c is the time at which $|h(t)| \leq 0.01\ |h_r(t)|_{max}$ for all $t > t_c$.

For time-delay error, the normalized integral of the response from zero to the time of the maximum is a more sensitive measure (particularly on a digital computer) than the more obvious measure of delay time to the

maximum. Therefore, the following measure is used for time-delay error:

$$e_d = \frac{\int_0^{t_m} |h_l(t)|dt}{|h_l(t)|_{max}} - \frac{\int_0^{t_m} |h_r(t)|dt}{|h_r(t)|_{max}} . \qquad (3.9)$$

where t_m is the time when the reference pulse reaches its maximum. The reason for this choice of measurements will be discussed in the next section which describes the method of evaluation.

3.4 Evaluation

The evaluation phase is used to determine the direction and magnitude of the parameter adjustment necessary for cancellation of the zero by the pole. The procedures used for one, two, or three pairs of dipoles in the isolated region will be discussed in order.

a. One real pole in the isolated region:

If there is only one real pole of the closed-loop system in the isolated region, the significant response will be given by the cut-off network and the dipole formed by this real pole and the zero of the specification network. The possible pole-zero plots for the one dipole and the cut-off network are shown in Figure 12. When the zero is closer to the origin than the pole, the impulse response will be like that in Figure 13 b . The delay time for this response is less than that of the reference from the cut-off network alone, so the time-delay error defined in Equation 3.9 is positive. The parameter should then be adjusted so that the pole is moved to the right until the pulse-width error is minimum. When the pole is dominant, the impulse response will be like that shown in Figure 13 c . The time-delay error is negative so the parameter controlling the real pole should

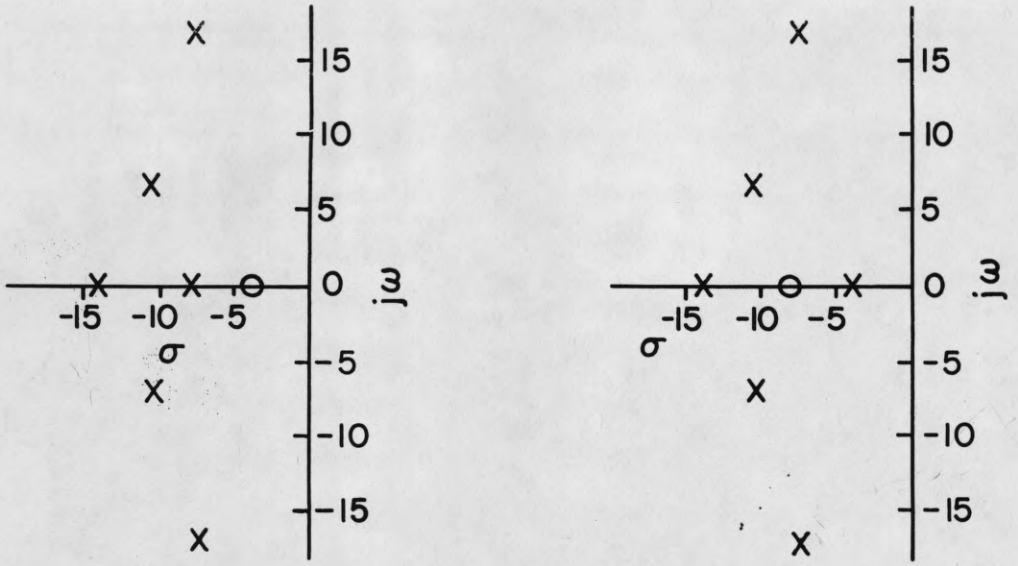
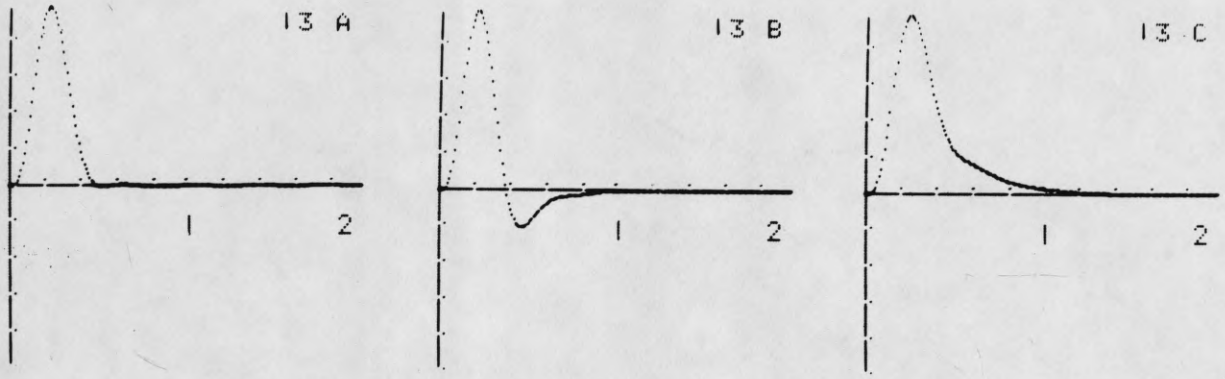


Figure 12. Pole-zero plots for a real dipole in the "isolated" region



a. cut-off network alone

b. zero dominant
$e_{\Delta t} = + 0.0072$
$e_d = + 0.0092$

c. pole dominant
$e_{\Delta t} = + 0.1030$
$e_d = - 0.0112$

Figure 13. Normalized impulse responses for a real dipole in the "isolated" region

be adjusted in the opposite direction for minimum pulse-width error.

b. Two complex poles in the isolated region:

When there are two complex conjugate poles in the isolated region, the significant response is given by the cut-off network and the two dipoles formed by the poles of the system and the zeros of the specification network. Two typical pole-zero plots are shown in Figure 14, and the corresponding impulse responses are shown in Figure 15. It is assumed that two parameters are available which, when varied, will cause the system poles to move along non-parallel loci. Again, the direction for adjustment of the parameter is determined by the time-delay error which is positive when the zeros are dominant and negative when the poles are dominant. Each parameter is adjusted until the pulse-width error is minimum so that at least three settings are needed for each parameter.

c. Two complex poles and a real pole in the isolated region:

If there are two complex conjugate poles and one real pole in the isolated region, the problem is more difficult because the poles of the set of complex dipoles may be dominant while the zeros of the real dipole are dominant or vice versa. There should be three variable parameters in this situation. The simplest procedure is to ignore the time-delay error and adjust the parameters, one at a time, for minimum pulse-width error. This procedure will require more steps to obtain the proper adjustment than the previous cases because the first step may be in the wrong direction. The time-delay error could also be used in determining the direction of the first step with the reservation that, if the first step did not reduce the pulse-width error, the adjustment would be switched to another parameter

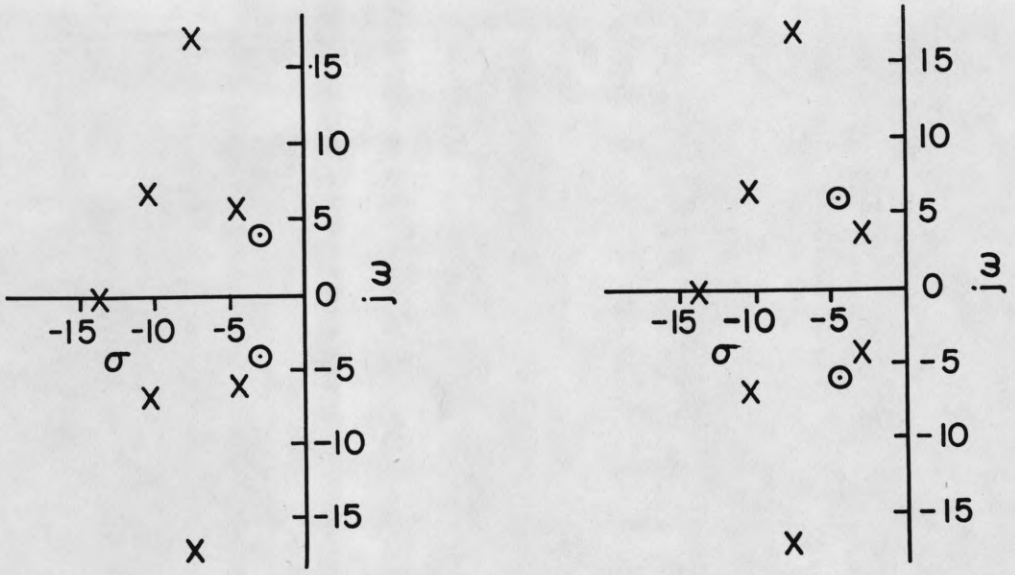
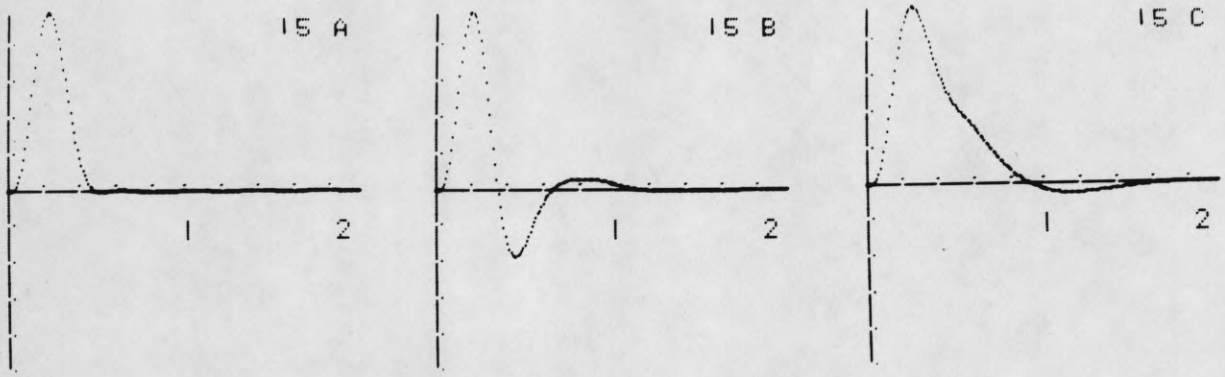


Figure 14. Typical pole-zero plots for a pair of complex dipoles in the "isolated" region



a. cut-off network alone

b. zeros dominant
$e_{\Delta t} = + 0.0386$
$e_d = + 0.0105$

c. poles dominant
$e_{\Delta t} = + 0.1730$
$e_d = - 0.0147$

Figure 15. Normalized impulse responses for a pair of complex dipoles in the "isolated" region

until the pulse-width error was reduced by the first step.

3.5 Parameter Adjustments

The adjustment requirements are very severe and make the method difficult to apply to many practical systems. Parameters must be available in the compensation network and the feedback network to give somewhat independent control over the poles which are to be changed to meet the specifications. It is difficult to insert parameters which affect one pair of poles and not all of the poles. The open-loop gain, for example, affects the locations of all the closed-loop poles as evidenced by the root locus plot for most closed-loop systems when gain is the variable. Parameters are usually available, however, which affect some poles much more than others. This is evidenced in the lengths of the various loci in the root locus plot. We will have to be content, therefore, with parameters to which one set of poles are more sensitive than the others with the hope of recycling the testing to obtain better cancellation. In a practical problem, we are usually content to get within a neighborhood of the specification zeros.

The adjustment phase will vary from system to system and can best be explained by use of an example. The method will be applied to a pitch-rate control system in the next chapter.

3.6 Step Function Inputs

In the previous discussion, the error measurement and evaluation phases were based on the impulse response. The cut-off and specification networks are linear, and it is assumed that the system is linear. The operations are still valid, therefore, for a step function input provided the derivative of the output of the cut-off network is used as the input to the error-measuring

apparatus. This operation is simple when the signal to noise ratio is high; for low signal to noise ratios, the differentiation will enhance the noise and reduce the feasibility of the modified model-reference system for step function inputs.

4. EXAMPLE: A PITCH-RATE CONTROL SYSTEM

The modified model-reference scheme will be applied to a practical pitch-rate control system in this chapter to indicate the procedure to be followed and to demonstrate the feasibility of the method.

4.1 Description of the System

A block diagram for the pitch-rate control system of a B-25 aircraft[17] is shown in Figure 16. This block diagram can be simplified to that of Figure 17. The closed-loop transfer function is:

$$T(s) = \frac{\dot{\theta}}{\delta_s}(s) = \frac{G}{1+GH} \tag{4.1}$$

where

$$G = \frac{K_I K_x K_a K_b K_c (s + a_c)}{s(s + a_b)(s^2 + 2\zeta_a \omega_a s + \omega_a^2)(s^2 + 2\zeta_c \omega_c s + \omega_c^2)} \tag{4.2}$$

$$H = \frac{K_1}{K_I}\left(s^2 + a_1 s + \frac{K_2 K_I}{K_1}\right) \tag{4.3}$$

Letting $K_x K_a K_b K_c = K$, we obtain:

$$T(s) = \frac{N(s)}{D(s)} \tag{4.4}$$

where

$$N(s) = K_I K(s + a_c) \tag{4.5}$$

$$D(s) = s(s + a_b)(s^2 + 2\zeta_a \omega_a s + \omega_a^2)(s^2 + 2\zeta_c \omega_c s + \omega_c^2)$$

$$+ K_1 K(s + a_c)(s^2 + a_1 s + K_2 K_I / K_1). \tag{4.6}$$

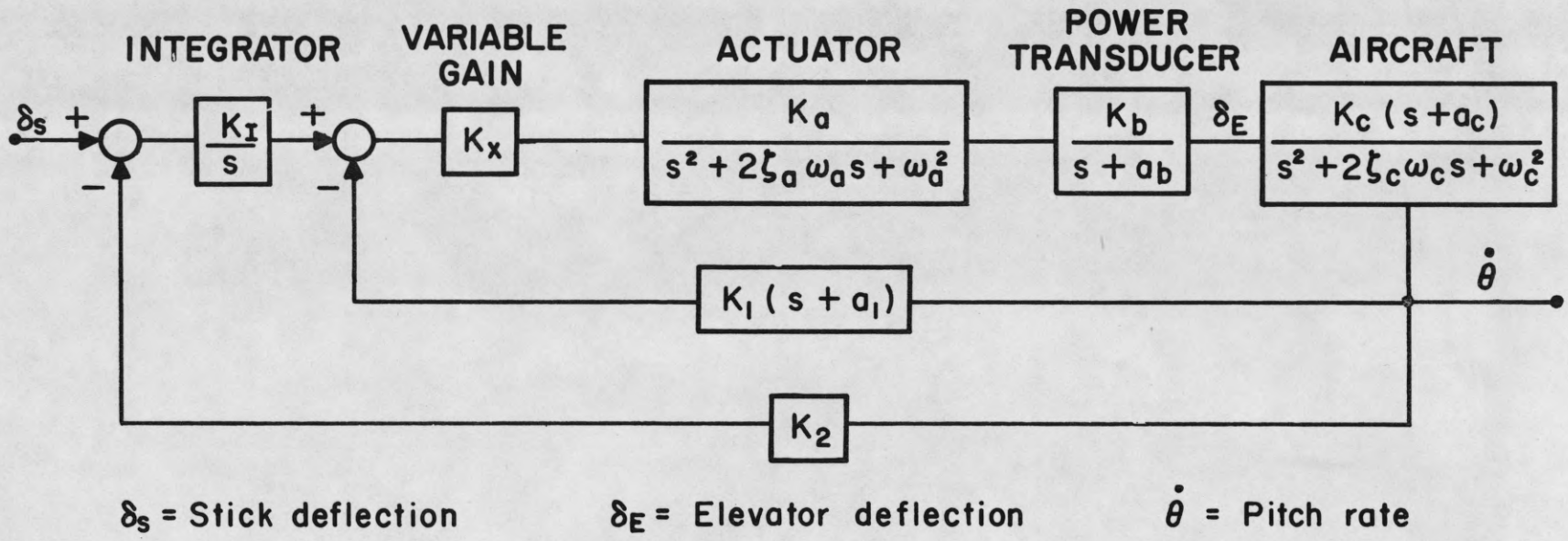
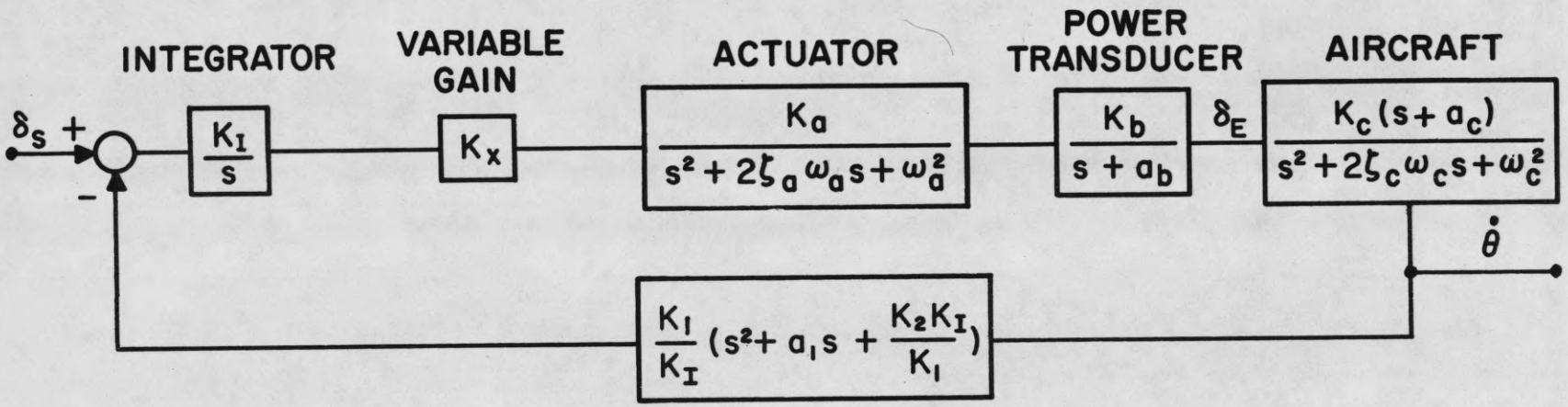


Figure 16. Block diagram of a pitch-rate control system for a B-25 aircraft



a_c = Aircraft path cut-off frequency
ζ_c = Aircraft short-period damping ratio
ω_c = Aircraft short-period resonant frequency
$a_1 = K_3/K_1$

K_c = Aircraft short-period gain
K_1 = Rate gyro gradient
K_2 = Angular accelerometer gradient
K_3 = Rate gyro gradient

Figure 17. Simplified block diagram for a pitch-rate control system

Typical values for the fixed parameters are:

$\omega_a = 50$ rad/sec. $\qquad a_1 = 20$ rad/sec.

$\zeta_a = 0.7$ $\qquad \frac{K_2 K_I}{K_1} = 200$ (rad/sec.)^2

$a_b = 15$ rad/sec.

It is assumed that the aircraft gain, K_c, the aircraft path cut-off frequency a_c, the aircraft frequency, ω_c, and the aircraft damping ratio, ζ_c, are variable.

The pitch-rate control problem handled by the General Electric Company[17] is described below. The amount of the aircraft parameter variation is indicated by the root locus plots for the two extremes of the flight regime. These root locus plots are shown in Figure 18 and Figure 19 with K, the open-loop gain, used as the variable parameter. Parameters in the compensation network and feedback network are to be varied so that the closed-loop system will remain invariant and give the desired response despite the undesirable changes in the aircraft transfer function. The desired operating regions are indicated by the squares on the root locus plots. The real pole near the origin can be ignored because it is effectively canceled by the zero for the values of gains involved.

This problem is adequately handled by the General Electric system because the desired operating region can be obtained for both extremes by varying the single parameter K_x, the gain of the compensation network. In other words, the specifications are such that the root locus plot is not significantly different in the specified region of operation for either extreme of flight conditions; thus, a simple gain change is adequate.

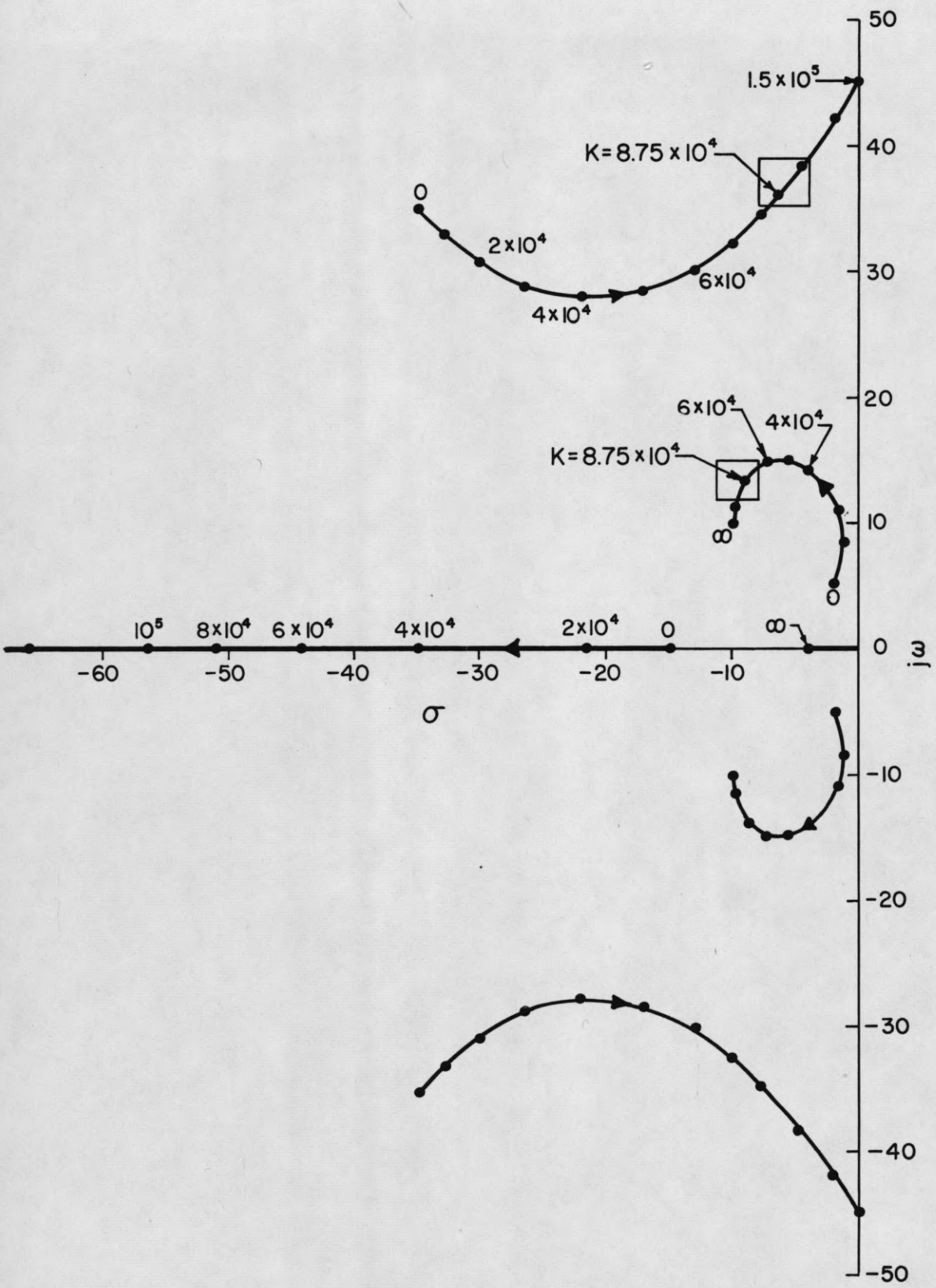


Figure 18. Root locus for one extreme of the flight regime

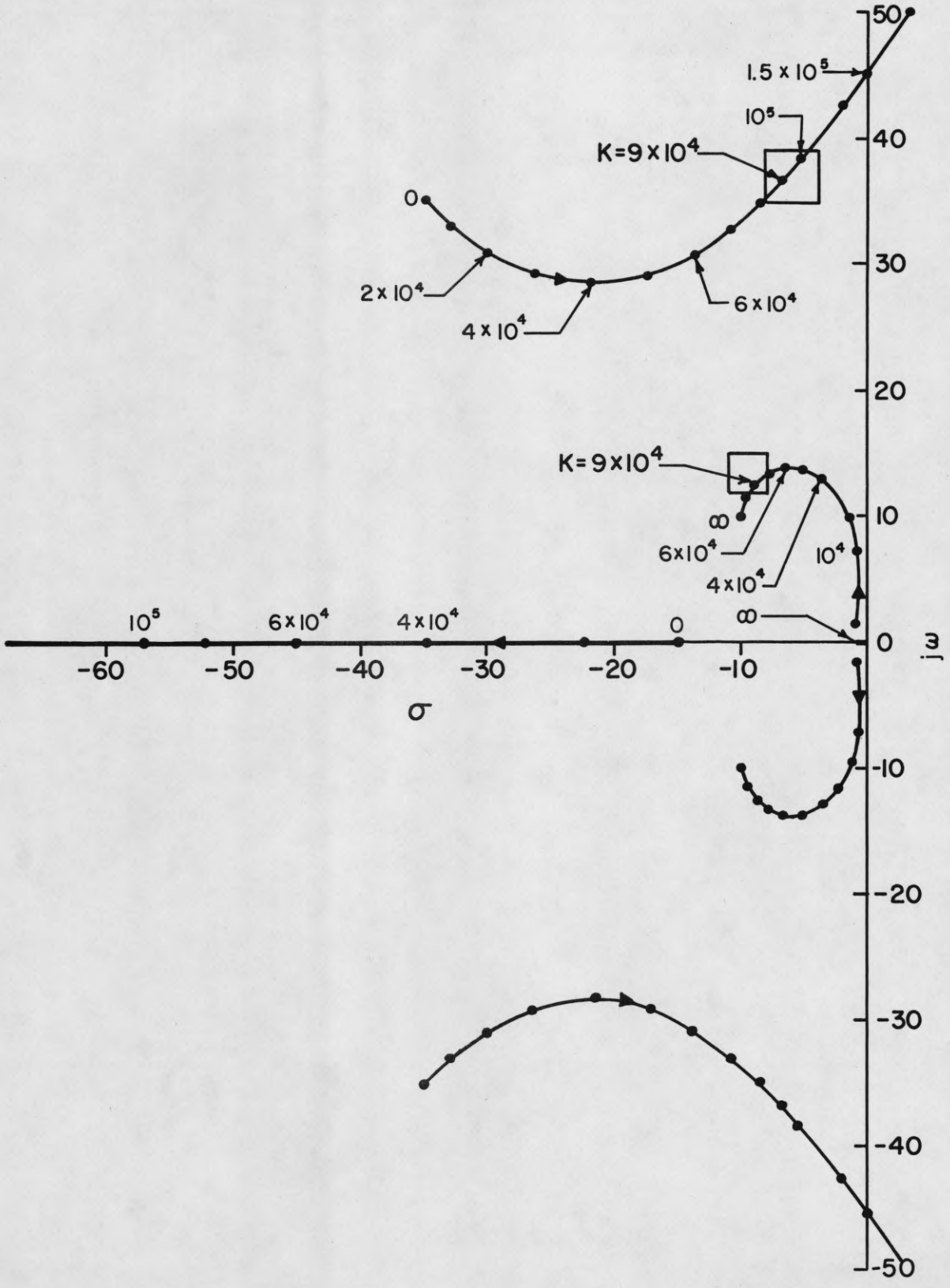


Figure 19. Root locus for the other extreme of the flight regime

4.2 Statement of the Problem

In order to increase the complexity of the problem, the specifications will be made more stringent by reducing the size of the region which meets the specifications, and the variation in the aircraft parameters will be made more extreme. The root locus diagram for this more complex situation is shown in Figure 20. The desired operating regions are indicated by the circles on this plot; the original specifications are shown by the squares. For this situation a simple gain change is not sufficient to obtain the desired operation. In fact, four parameters which give somewhat independent control over the two pair of complex poles are needed to meet a general specification. In this system there are three gains which can be readily used as variable parameters; namely, K_2, K_x, which varies the open-loop gain, and K_3, which varies a_1. Unless additional networks are added to the system, there is no variable parameter available which gives an effect independent of the three parameters mentioned above. Rather than alter the system to meet the requirements, the problem will be solved with only the three variable parameters, so the specifications may not be met exactly.

The root locus plot in Figure 20 was made with the open-loop gain as the variable parameter. Similar plots may be made with a_1 and K_2 as the variable parameters. The root locus diagram is a plot of the poles of the closed-loop transfer function versus a variable parameter. To obtain the root locus plot using the ILLIAC digital computer, the closed-loop transfer function is put into the form:

$$T(s) = \frac{p(s)}{q(s) + kr(s)} = \frac{N(s)}{D(s)} \qquad (4.7)$$

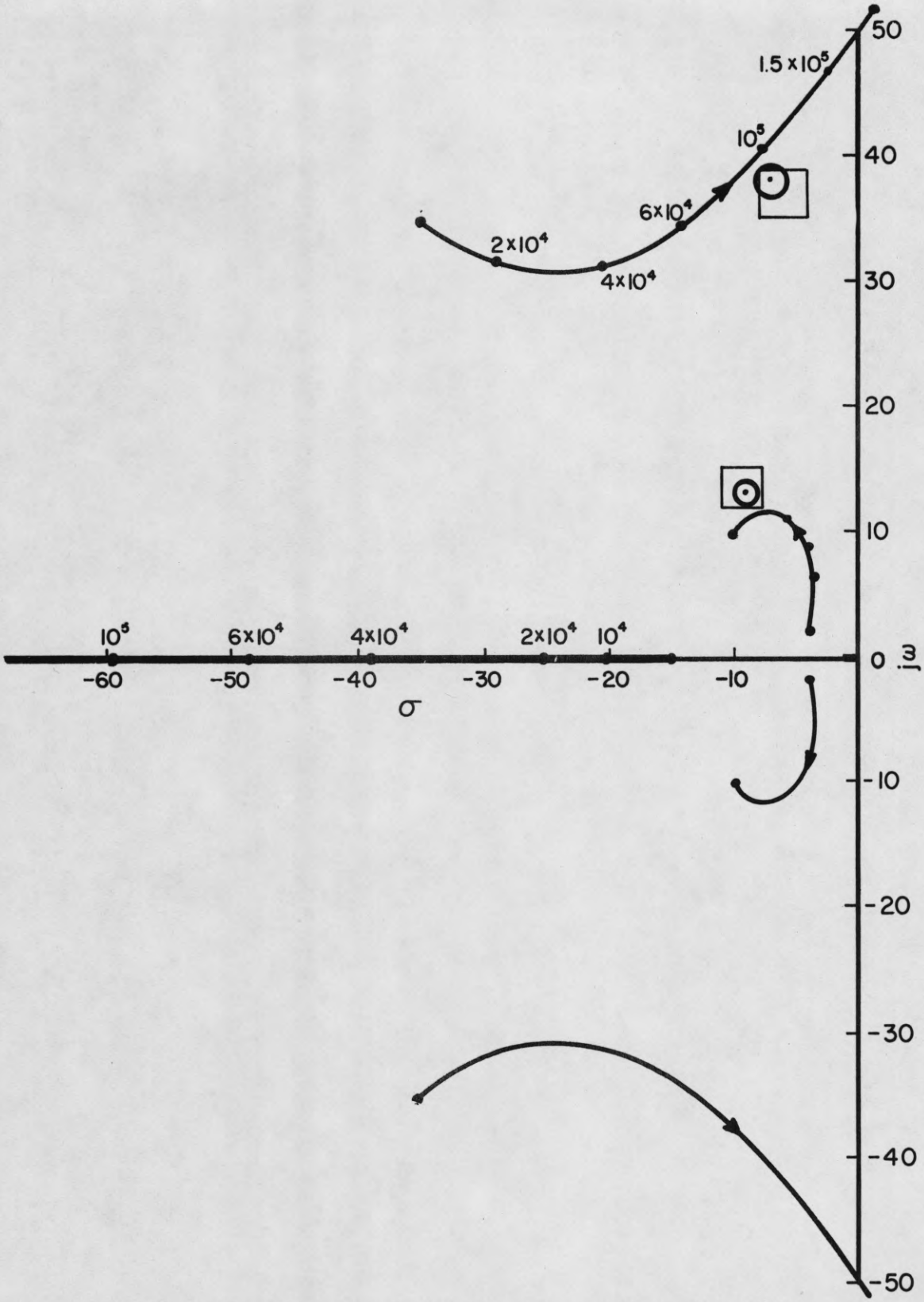


Figure 20. Root locus for the more complex situation

where k is the parameter to be varied. Then a plot is obtained of the zeros of $q(s) + kr(s)$ versus k.

With K_x as the variable parameter and the other gains fixed, we let $KK_1 = K_x K_a K_b K_c K_1$ be the variable parameter for the root locus plot, so there is a scale change between K_x and KK_1. The denominator becomes:

$$D(s) = [s(s + a_b)(s^2 + 2\zeta_a \omega_a s + \omega_a^2)(s^2 + 2\zeta_c \omega_c s + \omega_c^2)]$$

$$+ KK_1[(s + a_c)(s^2 + a_1 s + K_2 K_I / K_1)] \qquad (4.8)$$

With K_3 as the variable parameter, we let $a_1 = K_3/K_1$ be the variable parameter for the root locus plot so that again there is a scale change. The equation solved for the root locus plot becomes:

$$D(s) = [s(s + a_b)(s^2 + 2\zeta_a \omega_a s + \omega_a^2)(s^2 + 2\zeta_c \omega_c s + \omega_c^2)$$

$$+ KK_1(s + a_c)(s^2 + K_2 K_I / K_1)] + a_1[KK_1(s^2 + a_c s)] \qquad (4.9)$$

With K_2 as the variable parameter, we let $K_2 K_1 / K_1$ be the variable parameter for the root locus plot so that the equation is put into the form:

$$D(s) = [s(s + a_b)(s^2 + 2\zeta_a \omega_a s + \omega_a^2)(s^2 + 2\zeta_c \omega_c s + \omega_c^2)$$

$$+ KK_1(s + a_c)(s^2 + a_1 s)] + \frac{K_2 K_I}{K_1}[KK_1(s + a_c)] \qquad (4.10)$$

By varying one of the parameters with the other two fixed at various values, we can obtain a mapping of the entire complex frequency plane. Some of the possible loci for this problem are shown in Figure 21. The numbers on the loci correspond to KK_1, a_1, and $K_2 K_I / K_1$ while the actual parameters varied

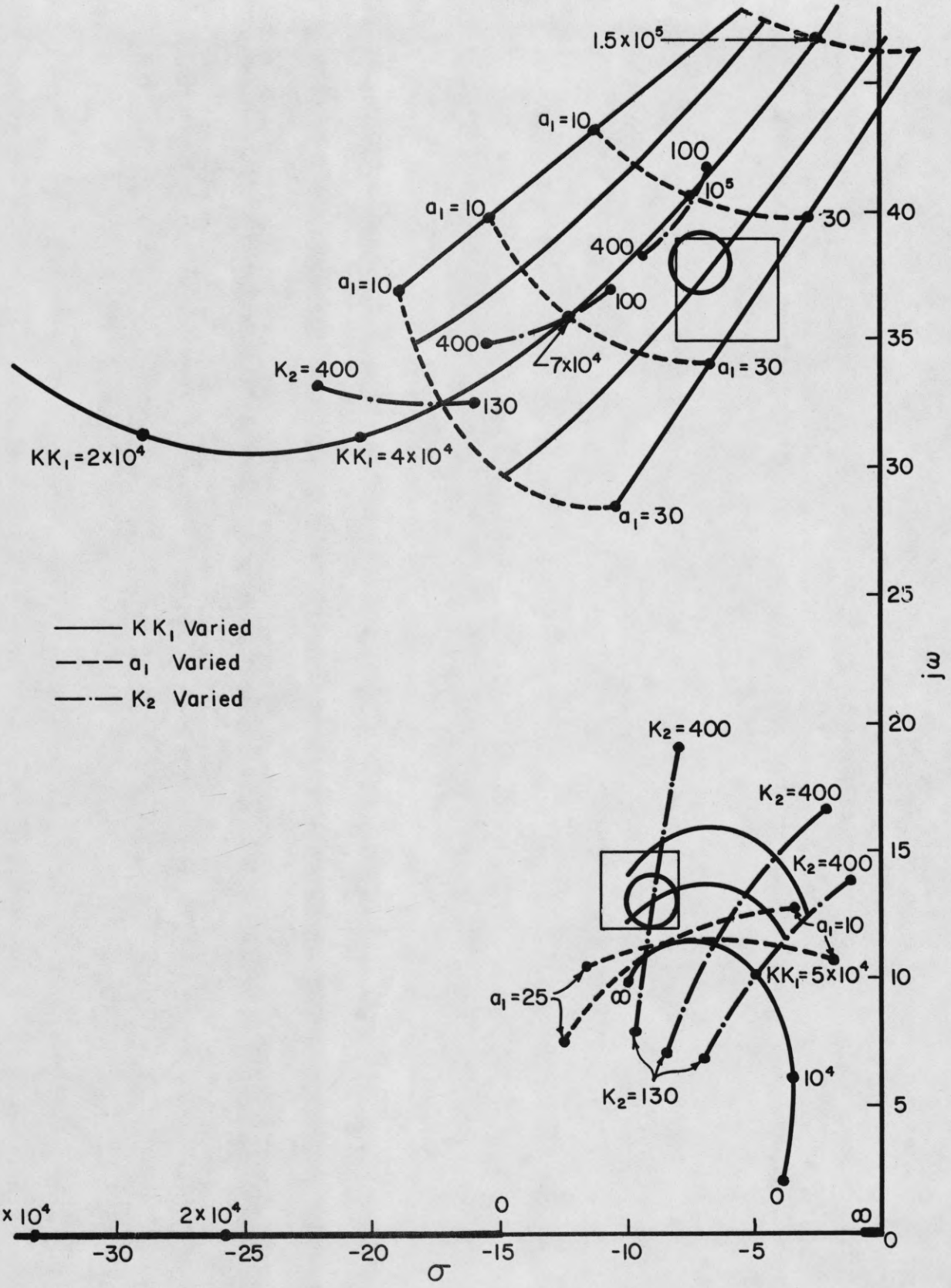


Figure 21. Root locus with KK_1, a_1, and K_2 as the variable parameters

in the system are K_x, K_3, and K_2, respectively.

Referring to Figure 21, it is noted that the parameter K_2 has a much stronger effect on the inner pair of complex poles than on the outer pair. Also, the parameter KK_1 has a more pronounced effect on the outer pair of complex poles. The parameter a_2, however, has a strong effect on both sets of poles. Thus, all the parameters do not give the independent control mentioned as a requirement in Section 3.5. Since problems of this sort will arise in most practical systems, it will be more practical to work with this set of parameters than to create an additional parameter to obtain independent control.

The problem to be handled by the modified model-reference scheme is as follows: The aircraft gain, K_c, is unknown while the other aircraft parameters are fixed at the values used for the root loci in Figures 20 and 21; namely,

$$a_c = 1 + j0$$

$$p_c = 4 \pm j2$$

The modified model-reference system is to vary the three parameters K_x, K_3, and K_2 until the closed-loop poles are within the specified region indicated by the circles in Figures 20 and 21.

Two different cut-off networks will be used as denoted by the poles in Figure 22. The innermost cut-off network is used when adjusting the inner complex poles until they are close to the specification of $s = -9 \pm j13$. The cut-off network is then switched to the outer one for adjusting the outer complex poles until they are close to the specification of $s = -7 \pm j38$.

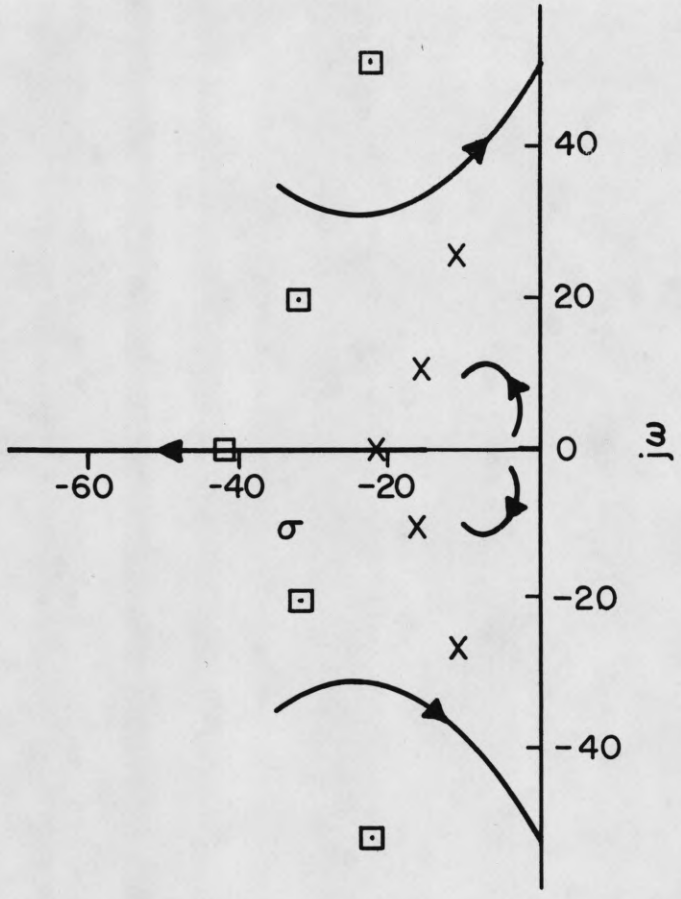
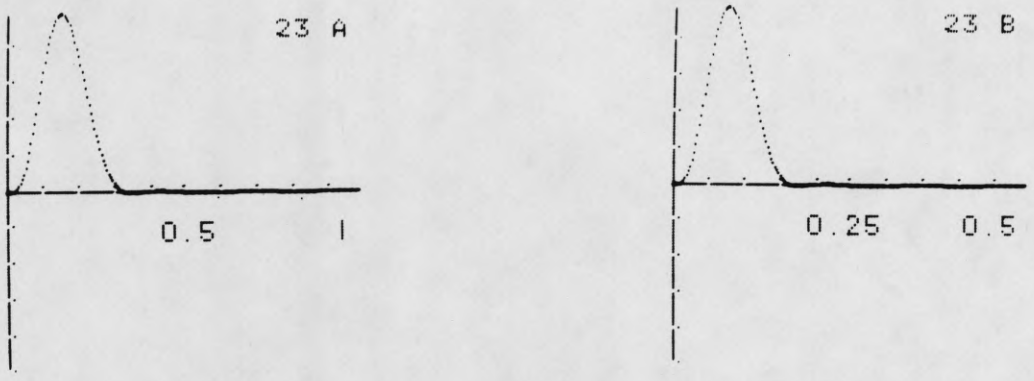


Figure 22. Poles of cut-off networks plus root locus of the system



a. inner cut-off network

b. outer cut-off network

Figure 23. Normalized impulse responses of the cut-off networks

The corresponding impulse responses that will be used as references are shown in Figure 23. The specification network should have the pole-zero plot shown in Figure 24. Since an all-zero transfer function is not practically realizable, the specification network and cutoff network in cascade with it are combined into a single network with the resultant transfer function having the pole-zero plot shown in Figure 25. This transfer function is still not practically realizable since the number of zeros is as great as the number of poles; therefore, at least a seventh order cut-off network should be used. An alternative is to use the fifth-order network and place two additional poles far out in the s plane so that they have negligible effect on the shape of the filter response. The second method was used in this problem to obtain a seventh-order cut-off network.

4.3 Experimental Results

The experimental work was done using the ILLIAC digital computer. The initial locations of the poles of the closed-loop system before the self-adaptive process began are shown as the points L-1 in Figure 26. These actual locations were not known to the testing equipment, and it operated completely on the basis of the pulse-width error and time-delay error. The direction of the proposed adjustment was obtained from the time-delay error; the parameter was then adjusted for minimum pulse-width error. This method of adjustment required at least three points for each parameter. A history of the steps taken is shown in Figure 26, and the interesting areas expanded in Figure 27.

The points L-1 thru L-7 were calculated with the cut-off network in the inner position, the object being to approximately cancel the inner pair of

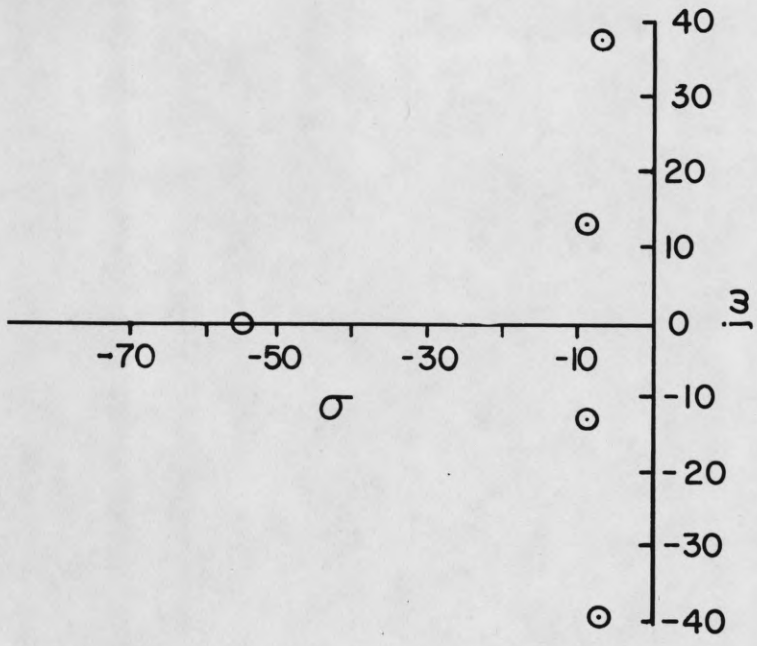


Figure 24. Pole-zero plot for the specification network

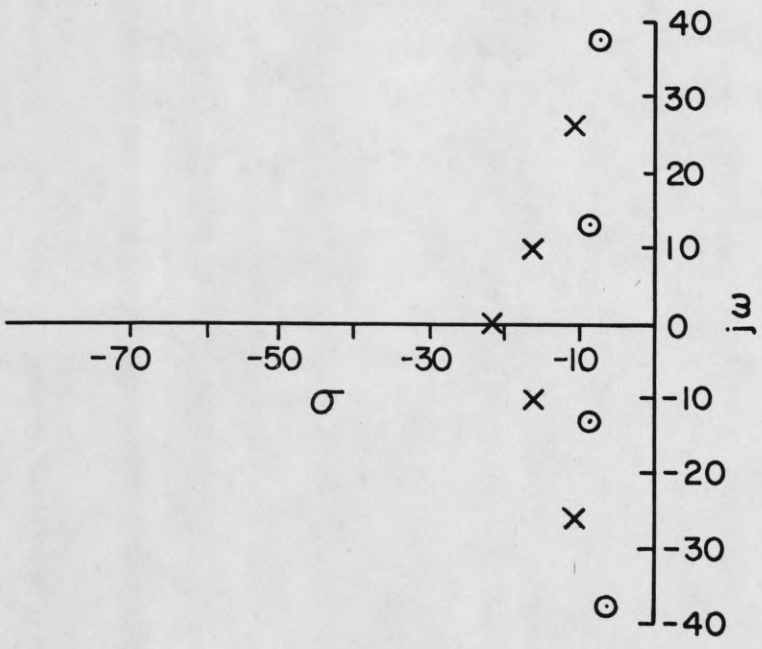


Figure 25. Pole-zero plot for the specification network and the inner cut-off network in cascade

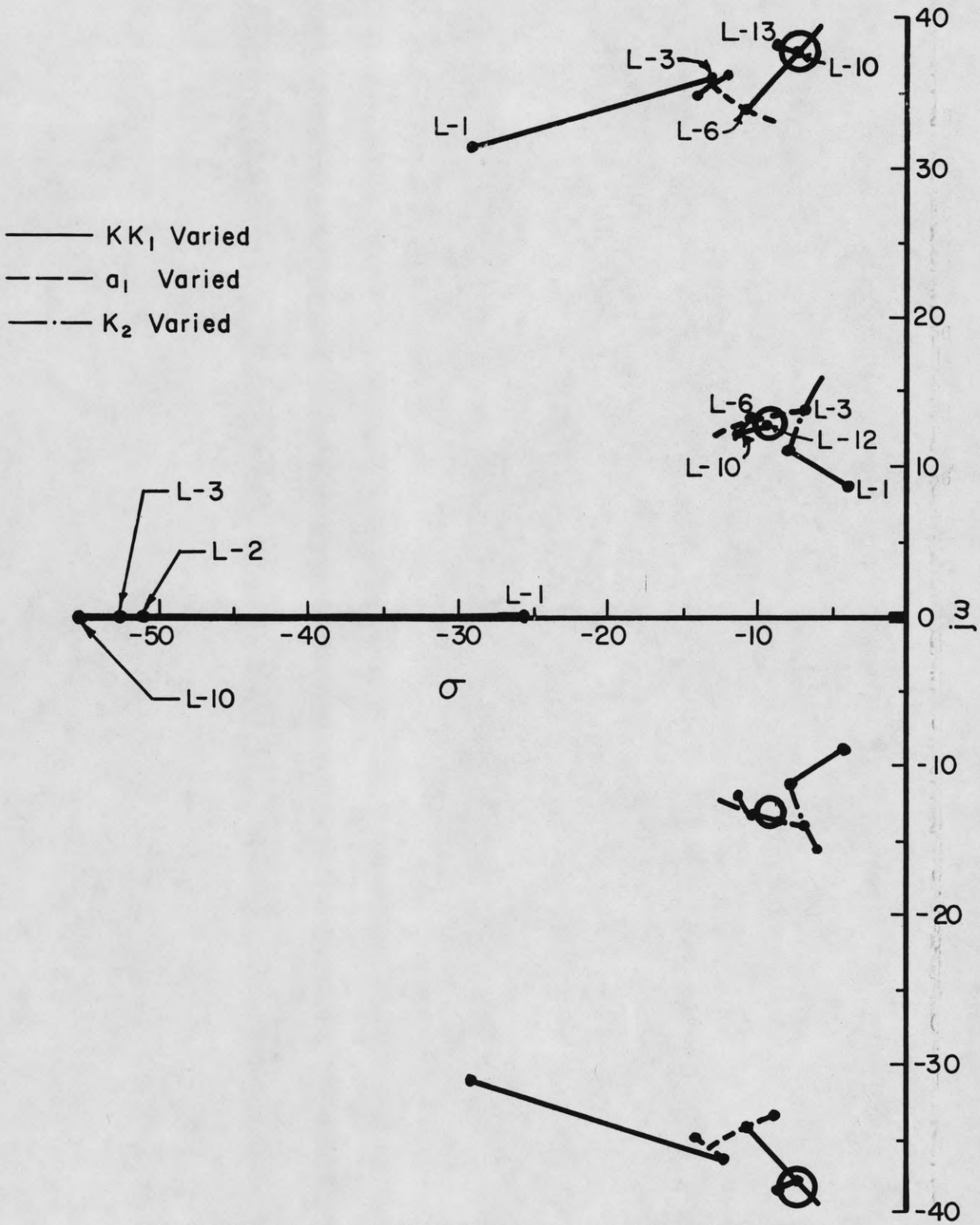


Figure 26. Locus followed by poles during adjustment

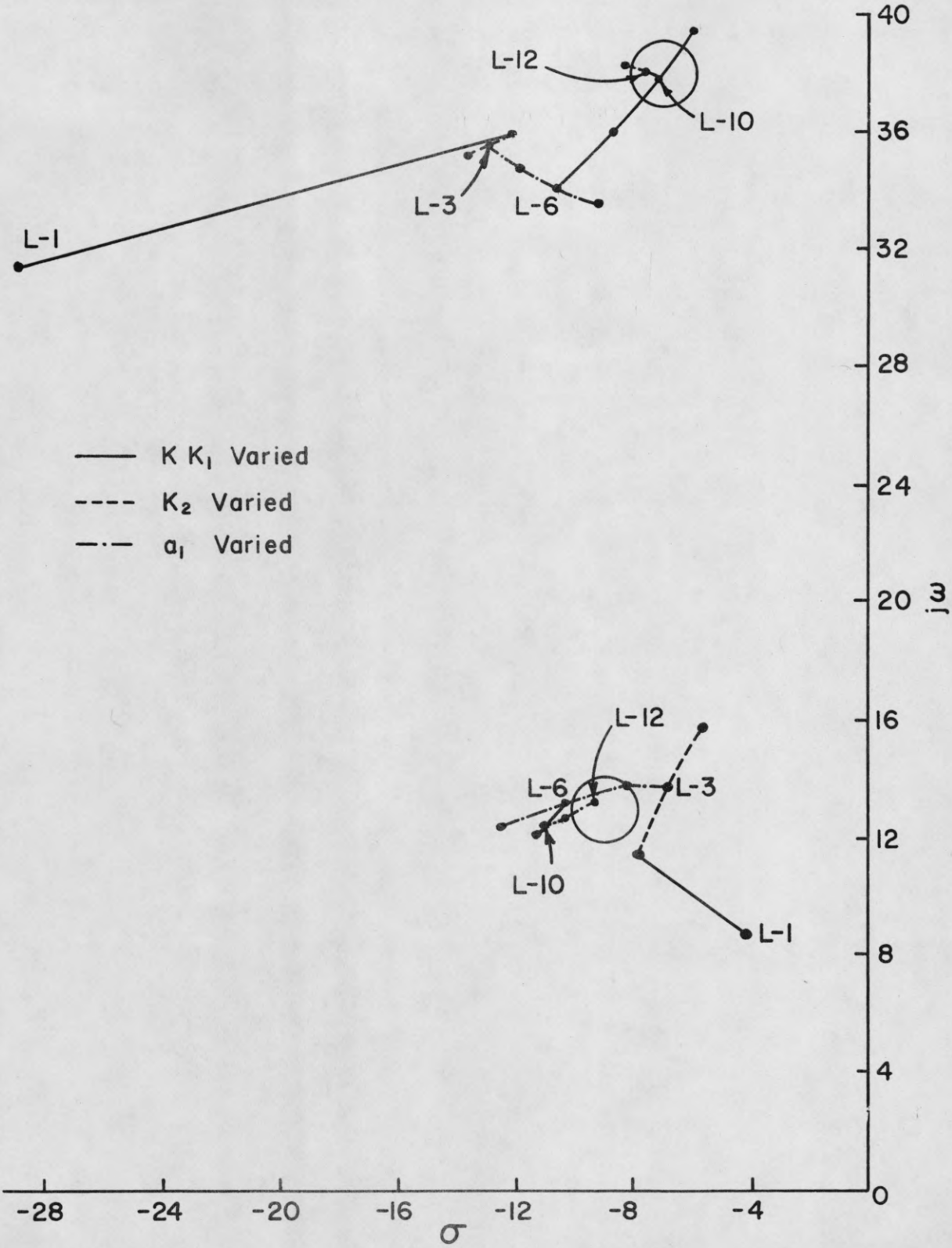


Figure 27. Expanded view of locus followed by poles during adjustment

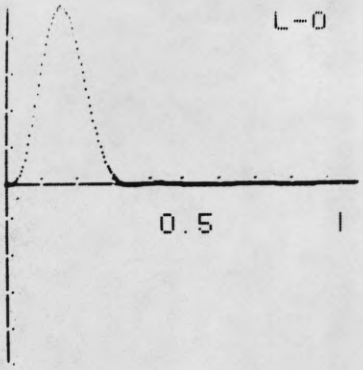
specification zeros. The outer cut-off network was then substituted, and points L-8 thru L-11 were used to adjust for cancellation of the outer pair of specification zeros. At this point the cut-off network was switched to the inner position and points L-12 thru L-14 were used as fine adjustments for cancellation of the inner specification zeros.

The open-loop gain was first increased in proceeding from L-1 to L-2. The parameter K_2 was then increased for L-3 and L-4, with L-3 giving the minimum pulse-width error. For L-5, L-6, and L-7, the parameter a_1 was increased with L-6 giving the minimum pulse-width error. For L-8 the cut-off network was switched to the outer network. For L-9, L-10, and L-11, the open-loop gain was again increased with L-10 giving the minimum pulse-width error. For L-12, L-13, and L-14, the cut-off network was switched back to the inner network and the parameter a_1 decreased with L-13 giving the minimum pulse-width error. This step was needed because both sets of poles were sensitive to the parameters, and the adjustments which improved the cancellation of one set of zeros degraded the cancellation of the other set. The computer calculations were stopped after L-14, but in system operation the process would be expected to continue. Smaller steps would be taken in the adjustments thereby forcing the system closer and closer to the desired state. As the changes in the environment change the system parameters, the system will continuously evaluate and adapt to these changes.

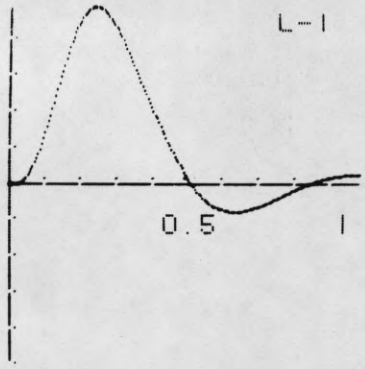
The impulse response at some of the key steps and the values of the pulse-width error and time-delay error are given in Figure 28. The steps chosen for the illustration are:

L-0 Reference for inner cut-off network

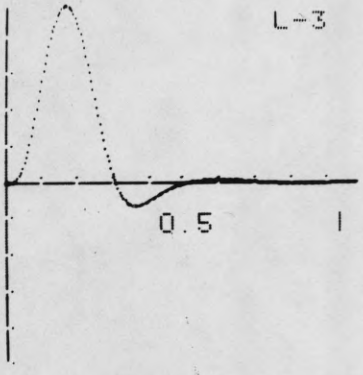
L-1 Initial settings



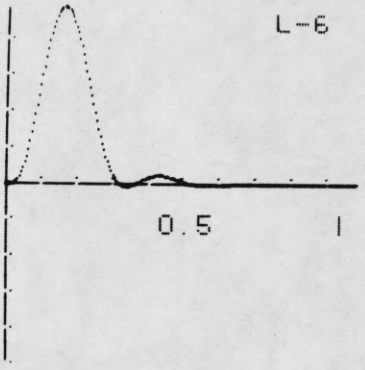
Reference



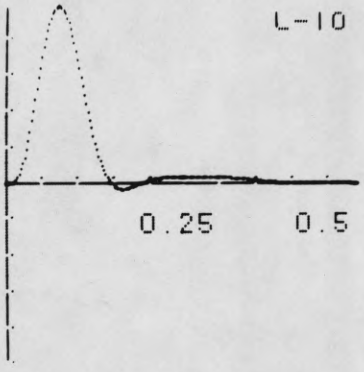
$e_{\Delta t}$ = + 0.1435
e_d = - 0.0344



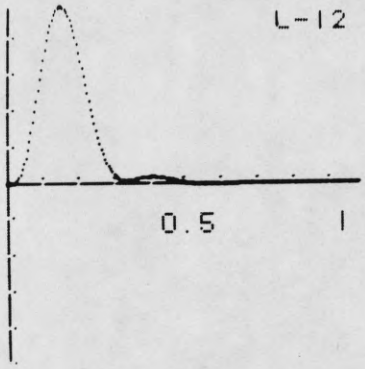
$e_{\Delta t}$ = + 0.0226
e_d = - 0.0082



$e_{\Delta t}$ = + 0.0086
e_d = - 0.0052



$e_{\Delta t}$ = + 0.0034
e_d = + 0.0012



$e_{\Delta t}$ = + 0.0041
e_d = + 0.0019

Figure 28. Normalized impulse responses at some of the key steps

L-3 Minimum $e_{\Delta t}$ for variation of K_2

L-6 Minimum $e_{\Delta t}$ for variation of a_1

L-10 Minimum $e_{\Delta t}$ for variation of KK_1

L-12 Minimum $e_{\Delta t}$ for variation of a_1

4.4 Convergence Time

The time required for convergence to the specification must be measured in terms of the number of applications of the input signal or "looks" at the impulse response. In the foregoing problem, the time for meeting the specifications from rather extreme initial conditions was that of twelve applications of the input signal. The inner specification zeros were approximately canceled after four adjustments or five "looks." The outer pair were approximately canceled after nine adjustments or ten "looks". In general, the minimum number of "looks" required are three for each parameter (the present value plus two adjustments) so that the minimum for this problem would be seven "looks". If the specifications are fairly loose or the change in aircraft parameters less severe, the specifications may be met before all six of the adjustments are made.

5. COMPARISON WITH OTHER SYSTEMS

Other possible methods of solving the aircraft pitch-rate control problem described in Chapter 4 will be discussed in this chapter, and a comparison will be made between these methods and the modified model-reference method. As mentioned in Chapter 4, we desire to maintain the specified system response despite changes in K_c, the aircraft gain, a_c, the aircraft path cutoff frequency, ω_c, the aircraft frequency, and ζ_c, the aircraft damping ratio. The modified model-reference method will handle this problem, but the need for such an elaborate scheme should be justified.

The simplest approach to the problem is to use an open-loop system. With this type of system, the sensitivity of the transmission function to changes in the aircraft gain is

$$S^T_{K_c} = 1 \tag{5.1}$$

A five percent change in K_c results in a five percent change in transmission gain. The sensitivity problem, then, forces us to use some type of feedback.

5.1 Conventional Feedback Control System

The simplest arrangement employing feedback is the conventional feedback control system in which the output is fed back to the input through a feedback network. A block diagram of the conventional system is shown in Figure 29. The sensitivity of the transmission function to K_c becomes

$$S^T_{K_c} = \frac{1}{1 + KK_c\, GH} \tag{5.2}$$

where KK_c is the transmission gain.

If the quantity KK_cGH can be made large for the frequencies of interest,

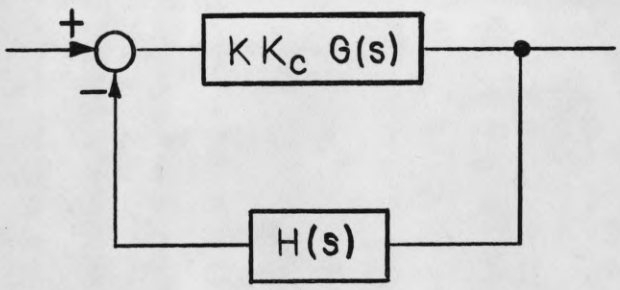


Figure 29. Block diagram of the conventional feedback control system

the sensitivity of the transmission function to changes in the aircraft gain can be reduced. A plot of the sensitivity of the transmission function of the pitch-rate control system to small single-parameter variations is given in Figure 30. For frequencies up to 6.5 radians per second, a change of one percent in the parameters will give less than a one-tenth percent change in the transmission function. At frequencies above 20 radians per second, the sensitivity to changes in K_c is greater than one, so the closed-loop system is worse than the open-loop system if these are the frequencies of interest.

For large parameter variation, the sensitivity function given in Equation 2.17 is valid only if the transmission function is a linear function of the parameter. The procedure for large multiple-parameter variations described in Section 2.4 will be applied to this problem. The variable parameters are first set at either extreme and the over-all transmission $T(j\omega)$, is calculated for all the combinations of parameter settings. The transmission $T_n(j\omega)$ is then calculated with all parameters set at their

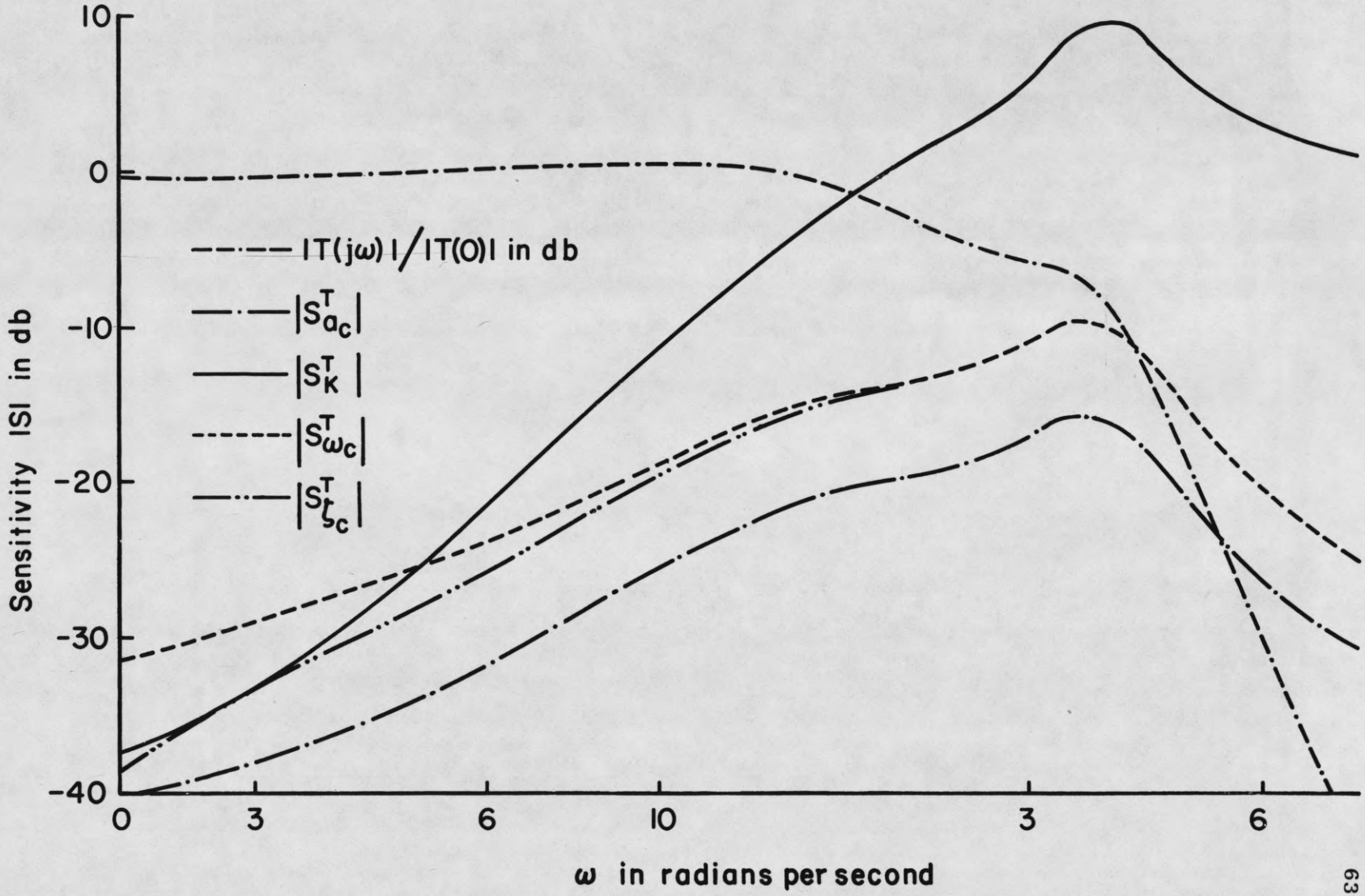


Figure 30. Plot of the sensitivity to small single-parameter variation for the pitch-rate control problem

normal value. The parameter settings are:

$(K_1K)_{min} = 7 \times 10^4$ $\quad$ $(K_1K)_{max} = 11 \times 10^4$ $\quad$ $(K_1K)_n = 9 \times 10^4$

$\zeta_{c_{min}} = 0.35$ $\quad$ $\zeta_{c_{max}} = 0.9$ $\quad$ $\zeta_{c_n} = 0.5$

$a_{c_{min}} = 1.0$ rad/sec. $\quad$ $a_{c_{max}} = 4.0$ rad/sec. $\quad$ $a_{c_n} = 2.0$ rad/sec.

ω_c is fixed at 4 rad/sec.

From the eight calculated values of $T(j\omega)$ at each frequency, $T_m(j\omega)$ is chosen as the one which deviates most from $T_n(j\omega)$. A plot of

$$\left| 20 \log \left| \frac{T_m(j\omega)}{T_n(j\omega)} \right| \right|$$

is given in Figure 31. The upper bound on the deviation may be greater than these values of $T_m(j\omega)$.

The results of the large multiple-parameter variations are similar to the sensitivity of $T(j\omega)$ to small variations in gain. The variation in $T(j\omega)$ is greatest at about 40 radians per second.

Whether a given sensitivity function is tolerable or not depends on the type of input signal applied to the system. The pitch-rate control system will have acceptable sensitivity properties for sinusoidal steady-state inputs below 6 radians per second. For step function inputs, however, the response at 40 radians per second is appreciable, and the sensitivity of the system transmission to parameter changes is too great to be acceptable.

Increasing the open-loop gain will not improve the sensitivity to changes in gain because the system becomes less stable. A completely different design using minor feedback loops will reduce the sensitivity to

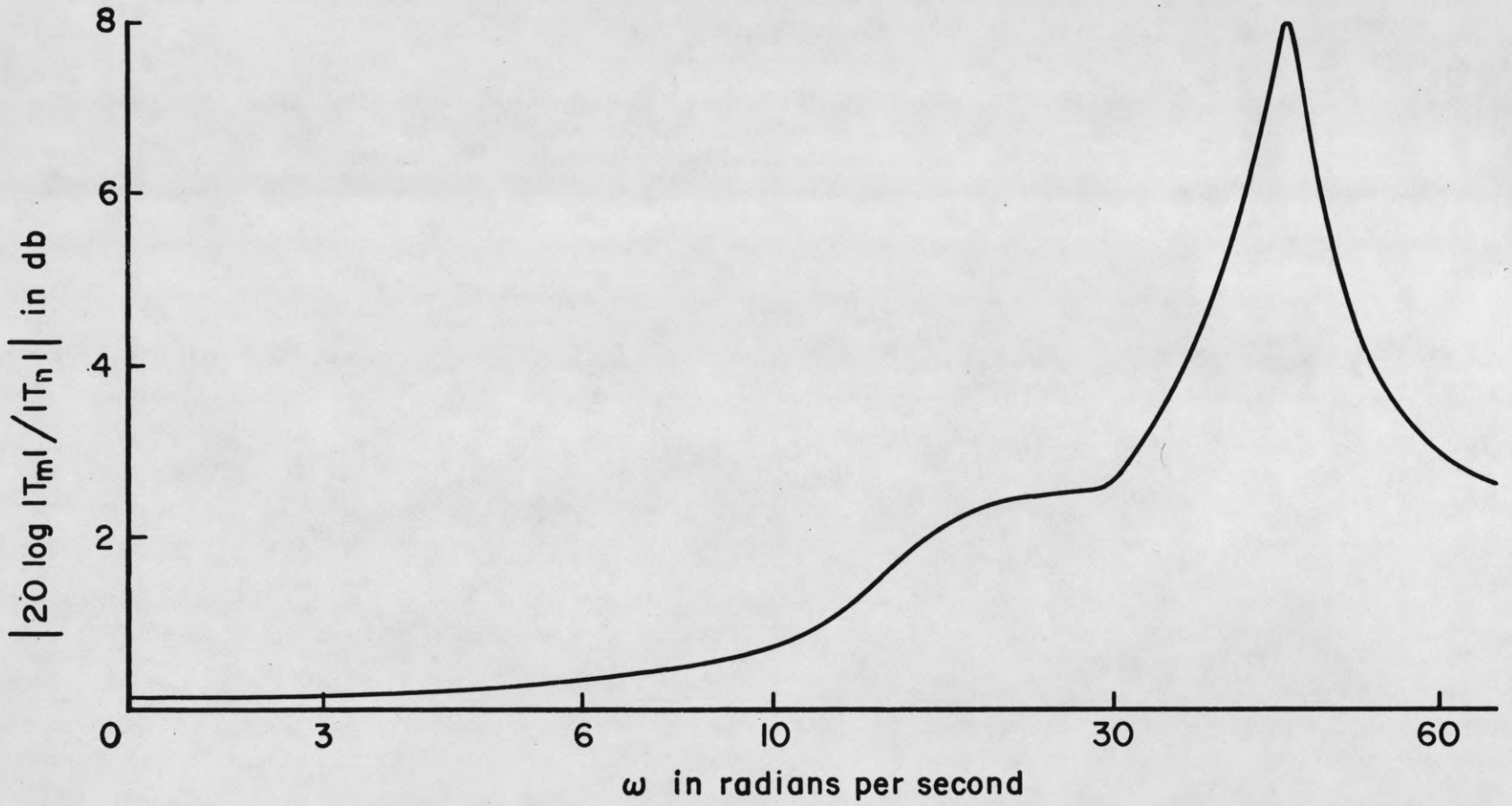


Figure 31. Plot of the transmission variation for large multiple-parameter variations

parameter changes. These parameter changes are so large, however, that it is doubtful that suitable performance could be achieved for step function inputs. The alternative, then, is to use an adaptive system that adjusts itself to compensate for the aircraft parameter changes. The existing methods of adaptive control are difficult to use because they were designed for second-order systems. The purpose of the following discussion is to show that only the model-reference scheme will adequately handle this high-order system.

5.2 The General Electric System

Using the General Electric approach to the problem[17], we work with the transfer function between the input and the point e on the block diagram shown in Figure 32. The transfer function of interest is:

$$\frac{E(s)}{\delta_s(s)} = \frac{\frac{K_I}{s}}{1 + \frac{K_1 K(s + a_c)(s^2 + a_1 s + K_2 K_I/K_1}{K_I\, s(s + a_b)(s^2 + 2\zeta_a \omega_a s + \omega_a{}^2)(s^2 + 2\zeta_c \omega_c s + \omega_c{}^2)}} \qquad (5.3)$$

The pole-zero plot for this transfer function is shown in Figure 33.

The frequency servo operates to adjust K_x for the proper undamped natural frequency, ω_n, of a second-order system. A prefilter must be inserted between the point e and the input to the frequency servo so that the input to the frequency servo is essentially the step response of a second-order system.

The impulse response at point e of p_b, p_b^*, z_b, and z_b^* is shown in Figure 34a. The impulse response of p_a, p_a^*, z_a, and z_a^* is shown in Figure 34b. The over-all response of the system with the pole-zero plot of Figure 33 is shown in Figure 34c. Because of the partial cancellation of

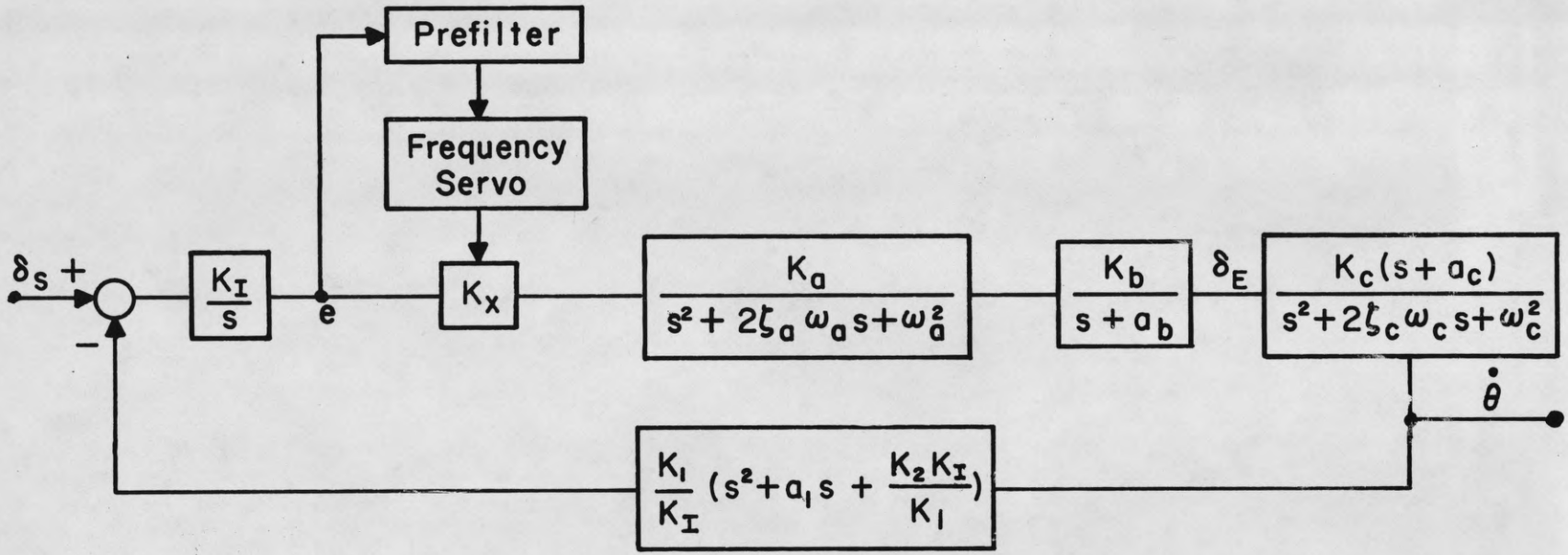


Figure 32. Block diagram of the General Electric system

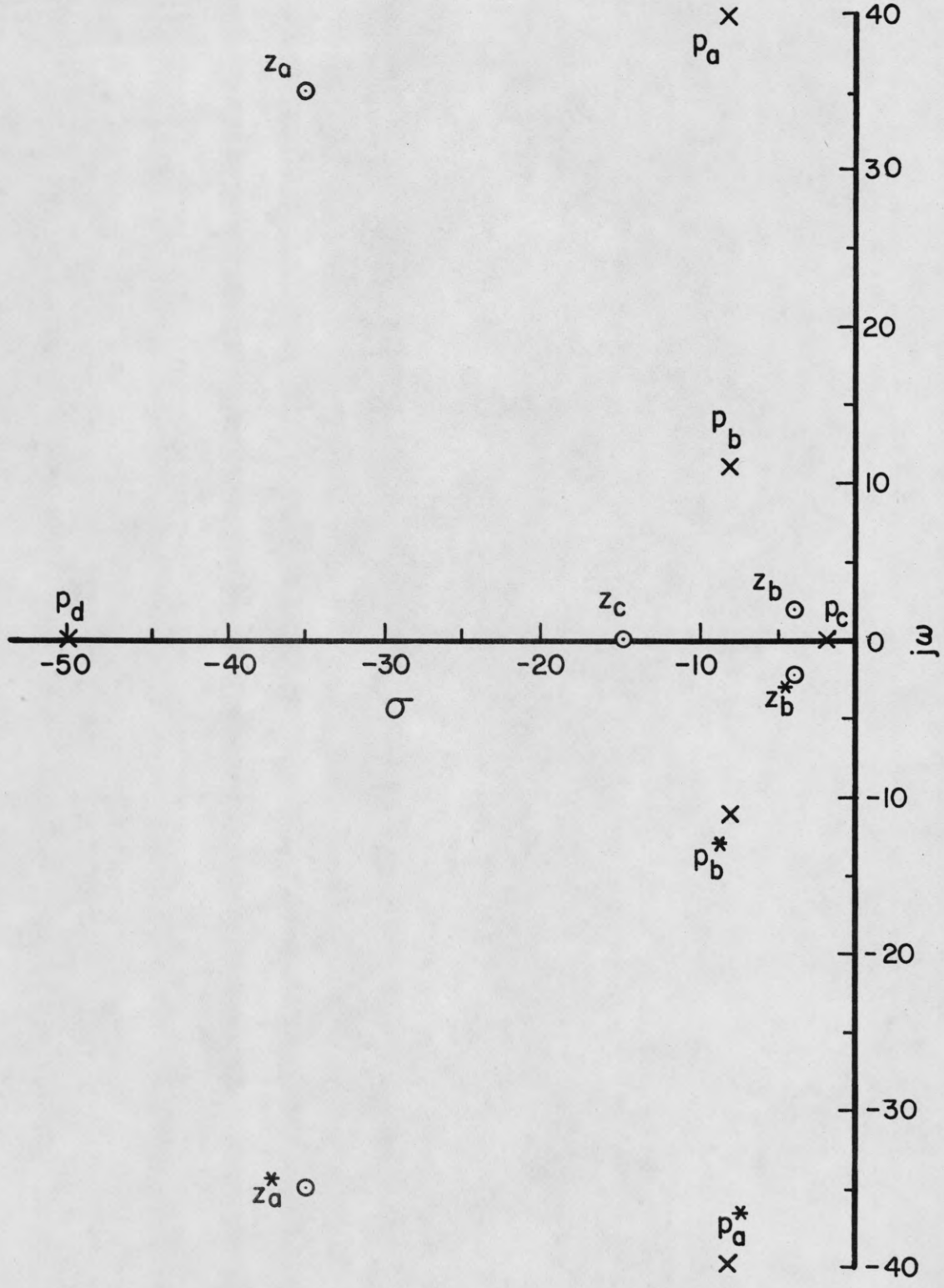
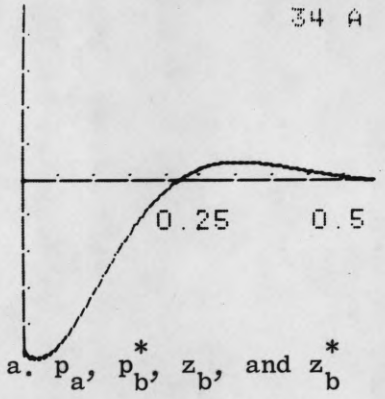
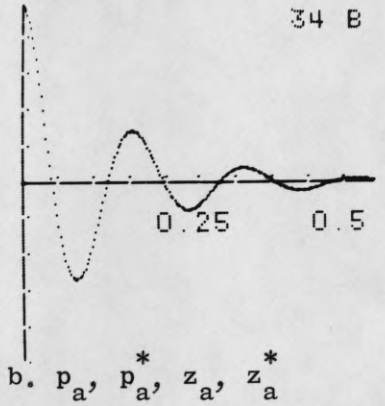


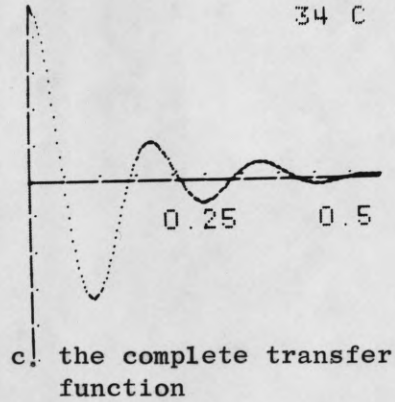
Figure 33. Pole-zero plot for the transfer function $E(s)/\delta_s(s)$



a. p_a, p_b^*, z_b, and z_b^*



b. p_a, p_a^*, z_a, z_a^*



c. the complete transfer function

Figure 34. Normalized impulse responses at point e.

the poles p_b and p_b^* by the zeros z_b and z_b^*, the poles p_a and p_a^* are dominant in the response at point e. The simplest prefiltering scheme would limit the response to that of the poles p_a and p_a^*. The poles dominant in the system response, $\dot{\theta}$, however, are the poles p_b and p_b^* because the zeros z_a, z_a^*, z_b, z_b^*, and z_c are not present in the transfer function $T(s) = \dot{\theta}(s)/\delta_s(s)$. The response, $\dot{\theta}$, to a step function input for the poles p_a and p_a^* is shown in Figure 35a; that for the poles p_b and p_b^* is shown in Figure 35b; and that for the complete transfer function in Figure 35c. These step responses show that all four poles affect the response, but the poles p_b and p_b^* are dominant in the system response. With this approach, then, we are adjusting so that the poles which have a lesser effect on the impulse response are correctly set.

No matter how the prefiltering is done, the over-all operation of the system is judged by the proper position of two complex poles. This method is not adequate to meet the specified response for the problem solved by the modified model-reference method because, regardless of the method of prefiltering, the desired operating conditions will not be maintained by monitoring only two of the poles.

5.3 The Aeronutronic System

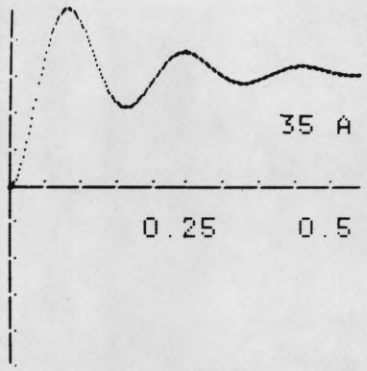
In the Aeronutronic system[2], the impulse response of the system is determined by injecting low-level noise at the system input and performing a cross-correlation of the input noise with the system output. For second-order systems, a figure of merit which is a measure of the system's damping ratio is calculated from the impulse response by comparing the positive and negative areas of the impulse response. The cross-correlation scheme

computes the impulse response for any order system. Once the impulse response has been computed, however, the calculation of the figures of merit becomes much more difficult for higher-order systems. The necessary parameter adjustments must be made from the information contained in these figures of merit. Future work may solve this problem of evaluating the impulse response of high-order systems, but the present methods will not handle the problem.

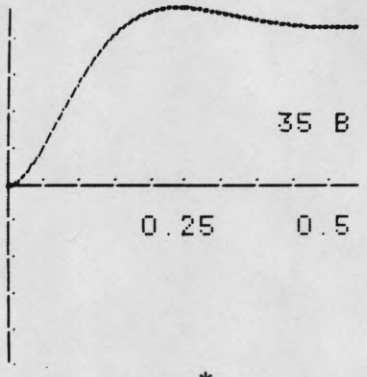
5.4 M.I.T. Model-Reference System

In the M.I.T. system, a model is constructed which gives the desired response. The output of the system under control is compared with that of the model and the parameters are adjusted for minimum error. This method handles the high-order system better than any of the other existing methods. The specification for a high-order system can be built into the model as easily as the second-order system specification. A difficulty arises in evaluating the error between the output of the model and that of the system. The method is kept simple when there are as many error criteria as adjustable parameters except that it is sometimes difficult to find an error criterion which is more sensitive to one parameter than to the others. The method will work with more adjustable parameters than error criteria, but it becomes almost as complex as the modified model-reference system because a computer is needed.

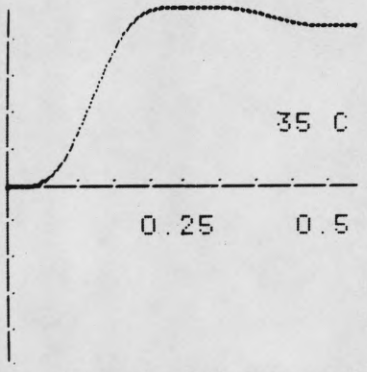
In the pitch-rate control problem, we need three error criteria, each of which is more sensitive to one parameter than the other two. Three likely criteria for step function inputs are the rise-time error, the integral of the absolute value of the error, and the integral of the error.



a. p_a and p_a^*



b. p_b and p_b^*



c. the complete transfer function

Figure 35. Normalized step responses at the output

The model-reference system can be made to work for the pitch-rate control problem. When more than three adjustable parameters are needed, however, the choice of a suitable error criteria becomes more difficult. The modified model-reference system has the advantage that both the step-by-step procedure followed and the two error measurements used are invariant regardless of the complexity of the problem.

One of the advantages of both the model-reference and modified model-reference system is that the measurement and evaluation schemes are not dependent on fixed positions of the pole-zero specifications. If certain missions require a change in transfer characteristics, these changes can be made in the specification network and the system will adjust itself to obtain the desired state. One does not, of course, have complete freedom in meeting an arbitrary specification, but a great deal of flexibility is available.

6. SUMMARY AND FURTHER PROBLEMS

6.1 Summary

A method of adaptive control has been developed that will adjust a high-order control system to meet a given specification. The specifications for desired operation were made in the form of the pole-zero locations of the transfer function of the desired system. The modified model-reference system used a special cut-off network whose impulse response was used as a reference and a specification network whose transfer function was the inverse of the model of the desired system. The cut-off networks were designed so that their impulse response had minimum variance about the time of the maximum. For step function inputs, the output of the cut-off networks was differentiated so that it was an impulse response. This reference output was shown to be dispersed when the transfer function of the control system did not meet the specifications. The time-delay error and pulse-width error, which are based on the shape of the output, were used as measures of dispersion. The necessary adjustments were found from the evaluation of these two error measurements.

The modified model-reference method was applied to a practical pitch-rate control system which was essentially a fourth-order system. In an experimental test of the theory, twelve applications of the input signal were required to meet stringent specifications from extreme initial conditions. The feasibility of other methods of handling the pitch-rate control problem was also discussed. The need for adaptive control was demonstrated and it was shown that only the model-reference schemes had the ability to adequately handle the fourth-order pitch-rate control problem.

6.2 Further Problems

The methods of adaptive control which incorporate a model of the desired state of the control system have the capability of an additional step in adaptive control. Rather than use a fixed model, the system could use a model which changes with the demands placed on the system. There are then, two stages of adaptation: the system adjusts itself to correspond to the model, and the model adjusts itself to give the desired response for each input. The second stage of adaptation presents the problem of synthesizing a time-variant model which will meet previously assigned input-output relationships. Each different input signal might require a new model.

Another unsolved problem that has application in adaptive control systems is a method of system synthesis which will result in parameters that give independent control over specific poles. It is desirable to have parameters to which all poles except the designated ones have zero sensitivity. This synthesis problem is particularly difficult because many of the parts of the system are fixed from other considerations, so the designer does not have complete freedom in his synthesis procedure.

Additional work is needed in determining the sensitivity for large multiple-parameter variations. The problem is to find a realistic upper bound on the variation of the transmission function without inordinate calculations.

Considerable effort is being expended toward the identification of the state of a system by calculation of its impulse response. The problem of converting the information contained in the impulse response to more useful s-plane information has only been solved for the second-order system. Additional work is required to make this method of identification useful for high-

order systems.

6.3 Conclusions

When problems arise which require an adaptive control system for which the second-order approximation is not valid, the model-reference system and the modified model-reference system are about the only methods that will adequately handle the problem. The advantages of the modified model-reference system over the model-reference system are that the modified scheme requires only two error measurements, and both the measurements to be made and the procedure to be followed are invariant. The disadvantages are that the modified model-reference scheme is more complex and the input signal is limited to pulses or step functions.

BIBLIOGRAPHY

1. Anderson, G.W., Aseltine, J.A., Mancini, A.R. and Sarture, C.W., "A Self-Adjusting System for Optimum Dynamic Performance," 1958 I.R.E. National Convention Record, Part 4, pp. 182-190.

2. Aseltine, J.A., Mancini, A.R., and Sarture, C.W., "A Survey of Adaptive Control Systems," I.R.E. Transactions on Automatic Control, Vol. PGAC-6, pp. 102-108, December 1958.

3. Bangert, J.T., "Practical Applications of Time Domain Theory," 1959 I.R.E. WESCON Convention Record, Part 2, Circuit Theory, pp. 29-38, August 18-21.

4. Bode, H.W., Network Analysis and Feedback Amplifier Design, D. Van Nostrand Co., Inc., 1945.

5. Brillouin, M.L., "Propagation des ondes electromagnetiques dans les milieux materials," Congres International D'Electricite, Vol. 2, pp. 739-788, 1932.

6. Brown, G.S., and Campbell, D.P., Principles of Servomechanisms, John Wiley and Sons, Inc., 1948.

7. Brussolo, J.A., "Pole Determinations with Complex-Zero Inputs," I.R.E. Transactions on Automatic Control, Vol. AC-4, pp. 150-166, November 1959.

8. Chen, K., "A Quick Method for Estimating Closed-Loop Poles of Control Systems," A.I.E.E. Transactions, Applications and Industry, pp. 80-87, May 1957.

9. Chestnut, H., and Mayer, R.W., Servomechanisms and Regulating System Design, Vol. 1, John Wiley and Sons, Inc., 1951.

10. Corbin, R.M., and Mishkin, E., "On the Measurement Problem in Adaptive Systems Utilizing Analog Computer Techniques," Polytechnic Institute of Brooklyn Research Report R-699-58 PIB-627, ASTIA AD211356, December 1958.

11. Cruz, J.B., "On the Realizability of Linear Differential Systems," I.R.E. Transactions on Circuit Theory, Vol CT-7, September 1960.

12. Di Toro, M.J., "Phase and Amplitude Distortion in Linear Networks," Proceedings of the I.R.E., Vol. 36, pp. 24-36, January 1948.

13. Exstrom, J.L., "The Impulse Response of an All-Pass Network Having Infinite-Order Phase Distortion," Proceedings of the I.R.E., Vol. 48, pp. 116-117, January 1960.

14. Eykhoff, P., "Adaptive and Optimalizing Control Systems," I.R.E. Transactions on Automatic Control, Vol. AC-5, pp. 148-151, June 1960.

BIBLIOGRAPHY (Continued)

15. Goodman, T.P., and Hillsley, R.H., "Continuous Measurement of Characteristics of Systems with Random Inputs: A Step Toward Self-Optimizing Control," A.S.M.E. Transactions, Vol. 80, pp. 1839-1848, November 1958.

16. Graham, D., and Lathrop, R.C., "The Synthesis of 'Optimum' Transient Response: Criteria and Standard Forms," A.I.E.E. Transactions, Applications and Industry, pp. 273-288, November 1953.

17. Gregory, P.C., "Proceedings of the Self Adaptive Flight Control Systems Symposium," Wright Air Development Center Technical Report, TR-59-49, ASTIA AD 209389, January 1959.

18. Guillemin, E.A., Communication Networks, Vol. 2, pp. 461-506, John Wiley and Sons, Inc., 1935.

19. Guillemin, E.A., Synthesis of Passive Networks, John Wiley and Sons, Inc., 1957.

20. Hakimi, S.L., and Cruz, J.B., "Measures of Sensitivity for Linear Systems with Large Multiple Parameter Variations," 1960 I.R.E. WESCON Convention Record, August 23-26.

21. Hartley, R.V.L., "Steady State Delay as Related to Aperiodic Signals," Bell System Technical Journal, Vol 20, pp. 222-234, April 1941.

22. Horowitz, I.M., "Fundamental Theory of Automatic Linear Feedback Control Systems," I.R.E. Transactions on Automatic Control, Vol AC-4, pp. 5-20, December 1959.

23. Kailath, T., "Sampling Models for Linear Time-Variant Filters," M.I.T. Research Lab. of Electronics Technical Report No. 352, May 25, 1959.

24. Kalman, R.E., "Design of a Self-Optimizing Control System," A.S.M.E. Transactions, Vol. 80, pp. 468-478, February 1958.

25. Lang, G., and Ham, J.M., "Conditional Feedback Systems--A New Approach to Feedback Control," A.I.E.E. Transactions, Applications and Industry, pp. 152-161, July 1955.

26. Lendaris, G.G. and Smith, O.J.M., "Complex-Zero Signal Generator for Rapid System Testing," A.I.E.E. Transactions, Applications and Industry, pp. 534-539, January 1959.

27. Lewis, J.B., "The Use of Nonlinear Feedback to Improve the Transient Response of a Servomechanism," A.I.E.E. Transactions, Applications and Industry, pp. 449-453, January 1953.

28. Mazer, W.H., "Specification of the Linear Feedback System Sensitivity Function," I.R.E. Transactions on Automatic Control, Vol. AC-5, pp. 85-93, June, 1960.

BIBLIOGRAPHY (Continued)

29. Mishkin, E., and Braun, L., Adaptive Control Systems, Department of Electrical Engineering, Polytechnic Institute of Brooklyn, Unpublished Notes, September 1959.

30. Mulligan, J.H., Jr., "The Effect of Pole and Zero Locations on the Transient Response of Linear Dynamic Systems," Proceedings of the I.R.E., Vol. 37, pp. 516-529, May 1949.

31. Osder, S.S., and Hutchinson, I.N., "Final Technical Report Feasibility Study Automatic Optimizing Stabilization System," Wright Air Development Center Technical Report TR-58-243, Part 1, ASTIA AD155576, June 1958.

32. Stratton, J.A., Electromagnetic Theory, pp. 330-340, McGraw-Hill Book Co., Inc., 1941.

33. Stromer, P.R., "Adaptive or Self-Optimizing Control Systems--A Bibliography," I.R.E. Transactions on Automatic Control, Vol. AC-4, pp. 65-68, May 1959.

34. Sussman, S.M., "Matched Filter Synthesis through Phase Distortion Networks," Rome Air Development Center Technical Report TR-58-6, ASTIA AD211356, pp. 74-88, June, 1957.

35. Truxal, J.G. and Horowitz, I.M., "Sensitivity Considerations in Active Network Synthesis," Proceedings of the Second Midwest Symposium on Circuit Theory, pp. 6-1 to 6-11, December 3-4, 1956.

36. Truxal, J.G., "Modern Network Theory and Its Application to Feedback Control," Proceedings of the Conference on Systems Engineering, Purdue University, pp. 79-104, July 25-27, 1955.

37. Truxal, J.G., Automatic Feedback Control System Synthesis, McGraw-Hill Book Co., Inc., 1955.

38. Woodward, P.M., Probability and Information Theory, with Applications to Radar, McGraw-Hill Book Co., Inc., 1953.

39. Whitaker, H.P., Yamron, J., and Kezer, A., "Design of Model-Reference Adaptive Control Systems for Aircraft," M.I.T. Instrumentation Laboratory Report R-164, September 1958.

40. Zadeh, L.A., "Frequency Analysis of Variable Networks," Proceedings of the I.R.E., Vol. 38, pp. 291-299, March 1950.