Chapter 10

Analysis and Design of Feedback Control Systems: Objectives and Methods

10.1 INTRODUCTION

The basic concepts, mathematical tools, and properties of feedback control systems have been presented in the first nine chapters. Attention is now focused on our major goal: *analysis and design* of feedback control systems.

The methods presented in the next eight chapters are linear techniques, applicable to linear models. However, under appropriate circumstances, one or more can also be used for some nonlinear control system problems, thereby generating approximate designs when the particular method is sufficiently robust. Techniques for solving control system problems represented by nonlinear models are introduced in Chapter 19.

This chapter is mainly devoted to making explicit the objectives and to describing briefly the methodology of analysis and design. It also includes one digital system design approach, in Section 10.8, that can be considered independently of the several approaches developed in subsequent chapters.

10.2 OBJECTIVES OF ANALYSIS

The three predominant objectives of feedback control systems analysis are the determination of the following system characteristics:

1. The degree or extent of system stability
2. The steady state performance
3. The transient performance

Knowing whether a system is absolutely stable or not is insufficient information for most purposes. If a system is stable, we usually want to know how close it is to being unstable. We need to determine its *relative stability*.

In Chapter 3 we learned that the complete solution of the equations describing a system may be split into two parts. The first, the steady state response, is that part of the complete solution which does not approach zero as time approaches infinity. The second, the transient response, is that part of the complete solution which approaches zero (or decays) as time approaches infinity. We shall soon see that there is a strong correlation between relative stability and transient response of feedback control systems.

10.3 METHODS OF ANALYSIS

The general procedure for analyzing a linear control system is the following:

1. Determine the equations or transfer function for each system component.
2. Choose a scheme for representing the system (block diagram or signal flow graph).
3. Formulate the system model by appropriately connecting the components (blocks, or nodes and branches).
4. Determine the system response characteristics.

Several methods are available for determining the response characteristics of linear systems. Direct solution of the system equations may be employed to find the steady state and transient solutions

(Chapters 3 and 4). This technique can be cumbersome for higher than second-order systems, and relative stability is difficult to study in the time-domain.

Four primarily graphical methods are available to the control system analyst which are simpler and more direct than time-domain methods for practical linear models of feedback control systems. They are:

1. The Root-Locus Method
2. Bode-Plot Representations
3. Nyquist Diagrams
4. Nichols Charts

The latter three are frequency-domain techniques. All four are considered in detail in Chapters 13, 15, 11, and 17, respectively.

10.4 DESIGN OBJECTIVES

The basic goal of control system design is meeting *performance specifications*. Performance specifications are the constraints put on system response characteristics. They may be stated in any number of ways. Generally they take two forms:

1. Frequency-domain specifications (pertinent quantities expressed as functions of frequency)
2. Time-domain specifications (in terms of time response)

The desired system characteristics may be prescribed in either or both of the above forms. In general, they specify three important properties of dynamic systems:

1. Speed of response
2. Relative stability
3. System accuracy or allowable error

Frequency-domain specifications for both continuous and discrete-time systems are often stated in one or more of the following seven ways. To maintain generality, we define a **unified open-loop frequency response function $GH(\omega)$**:

$$GH(\omega) \equiv \begin{cases} GH(j\omega) & \text{for continuous systems} \\ GH(e^{j\omega T}) & \text{for discrete-time systems} \end{cases} \tag{10.1}$$

1. Gain Margin

Gain margin, a measure of relative stability, is defined as the magnitude of the reciprocal of the open-loop transfer function, evaluated at the frequency ω_π at which the phase angle (see chapter 6) is $-180°$. That is,

$$\text{gain margin} \equiv \frac{1}{|GH(\omega_\pi)|} \tag{10.2}$$

where $\arg GH(\omega_\pi) = -180° = -\pi$ radians and ω_π is called the **phase crossover** frequency.

2. Phase Margin ϕ_{PM}

Phase margin ϕ_{PM}, a measure of relative stability, is defined as 180° plus the phase angle ϕ_1 of the open-loop transfer function at unity gain. That is,

$$\phi_{PM} \equiv [180 + \arg GH(\omega_1)] \text{ degrees} \tag{10.3}$$

where $|GH(\omega_1)| = 1$ and ω_1 is called the **gain crossover** frequency.

EXAMPLE 10.1. The gain and phase margins of a typical continuous-time feedback control system are illustrated in Fig. 10-1.

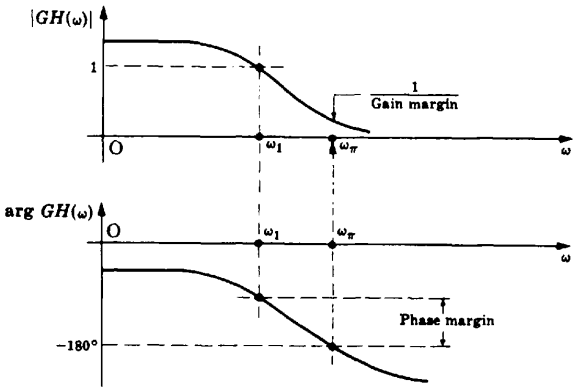


Fig. 10-1

3. Delay Time T_d

Delay time T_d, interpreted as a frequency-domain specification, is a measure of the speed of response, and is given by

$$T_d(\omega) = -\frac{d\gamma}{d\omega} \tag{10.4}$$

where $\gamma = \arg(C/R)$. The average value of $T_d(\omega)$ over the frequencies of interest is usually specified.

4. Bandwidth (BW)

Roughly speaking, the bandwidth of a system was defined in Chapter 1 as that range of frequencies over which the system responds satisfactorily.

Satisfactory performance is determined by the application and the characteristics of the particular system. For example, audio amplifiers are often compared on the basis of their bandwidth. An ideal high-fidelity audio amplifier has a *flat frequency response* from 20 to 20,000 Hz. That is, it has a passband or bandwidth of 19,980 Hz (usually rounded off to 20,000 Hz). Flat frequency response means that the *magnitude ratio* of output to input is essentially constant over the bandwidth. Hence signals in the audio spectrum are faithfully reproduced by a 20,000-Hz bandwidth amplifier. The magnitude ratio is the absolute value of the system frequency response function.

The frequency response of a high-fidelity audio amplifier is shown in Fig. 10-2. The magnitude ratio is 0.707 of, or approximately 3 db below, its maximum at the **cutoff frequencies**

$$f_{c1} = 20 \text{ Hz} \qquad f_{c2} = 20{,}000 \text{ Hz}$$

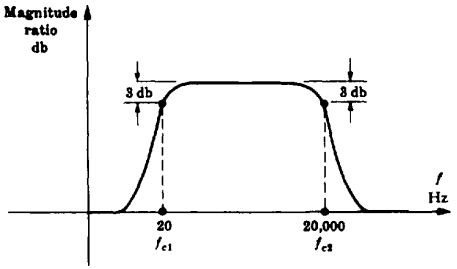


Fig. 10-2

"db" is the abbreviation for **decibel**, defined by the following equation:

$$\text{db} \equiv 20 \log_{10}(\text{magnitude ratio}) \tag{10.5}$$

Often the bandwidth of a system is defined as that range of frequencies over which the magnitude ratio does not differ by more than -3 db from is value at a specified frequency. But not always. In general, the precise meaning of bandwidth is made clear by the problem description. In any case, bandwidth is generally a measure of the speed of response of a system.

The gain crossover frequency ω_1 defined in Equation (*10.3*) is often a good approximation for the bandwidth of a closed-loop system.

The notion of signal sampling, and of *uniform sampling time* T, were introduced in Chapters 1 and 2 (especially in Section 2.4), for systems containing both discrete-time and continuous-time signals, and both types of elements, including samplers, hold devices and computers. The value of T is a design parameter for such systems and its choice is governed by both accuracy and cost considerations. The *sampling theorem* [9,10] provides an upper bound on T, by requiring the sampling rate to be at least twice that of the highest frequency component $f_{\max}$ of the sampled signal, that is, $T \le \dfrac{1}{2f_{\max}}$. In practice, we might use the cutoff frequency f_{c2} (as in Fig. 10-2) for $f_{\max}$, and a practical rule-of-thumb might be to choose T in the range $\dfrac{1}{10f_{c2}} \le T \le \dfrac{1}{6f_{c2}}$. Other design requirements, however, may require even smaller T values. On the other hand, the largest value of T consistent with the specifications usually yields the lowest cost for system components.

5. Cutoff Rate

The cutoff rate is the frequency rate at which the magnitude ratio decreases beyond the cutoff frequency ω_c. For example, the cutoff rate may be specified as 6 db/octave. An octave is a factor-of-two change in frequency.

6. Resonance Peak M_p

The resonance peak M_p, a measure of relative stability, is the maximum value of the magnitude of the closed-loop frequency response. That is,

$$M_p \equiv \max_{\omega} \left| \frac{C}{R} \right| \tag{10.6}$$

7. Resonant Frequency ω_p

The resonant frequency ω_p is the frequency at which M_p occurs.

EXAMPLE 10.2. The bandwidth BW, cutoff frequency ω_c, resonance peak M_p, and resonant frequency ω_p for an underdamped second-order continuous system are illustrated in Fig. 10-3.

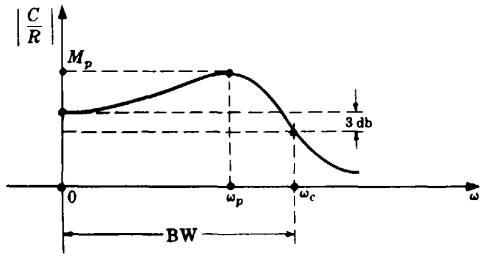


Fig. 10-3

Time-domain specifications are customarily defined in terms of unit step, ramp, and parabolic responses. Each response has a steady state and a transient component.

Steady state performance, in terms of steady state error, is a measure of system accuracy when a specific input is applied. Figures of merit for steady state performance are, for example, the error constants K_p, K_v, and K_a defined in Chapter 9.

Transient performance is often described in terms of the unit step function response. Typical specifications are:

1. Overshoot

The overshoot is the maximum difference between the transient and steady state solutions for a unit step input. It is a measure of relative stability and is often represented as a percentage of the final value of the output (steady state solution).

The following four specifications are measures of the speed of response.

2. Delay Time T_d

The delay time T_d, interpreted as a time-domain specification, is often defined as the time required for the response to a unit step input to reach 50% of its final value.

3. Rise Time T_r

The rise time T_r is customarily defined as the time required for the response to a unit step input to rise from 10 to 90 percent of its final value.

4. Settling Time T_s

The settling time T_s is most often defined as the time required for the response to a unit step input to reach and remain within a specified percentage (frequently 2 or 5%) of its final value.

5. Dominant Time Constant

The dominant time constant τ, an alternative measure for settling time, is often defined as the time constant associated with the term that dominates the transient response.

The dominant time constant is defined in terms of the exponentially decaying character of the transient response. For example, for first and second-order underdamped continuous systems, the transient terms have the form $Ae^{-\alpha t}$ and $Ae^{-\alpha t}\cos(\omega_d t + \phi)$, respectively ($\alpha > 0$). In each case, the decay is governed by $e^{-\alpha t}$. The time constant τ is defined as the time at which the exponent $-\alpha t = -1$, that is, when the exponential reaches 37% of its initial value. Hence $\tau = 1/\alpha$.

For continuous feedback control systems of order higher than two, the dominant time constant can sometimes be estimated from the time constant of an underdamped second-order system which approximates the higher system. Since

$$\tau \leq \frac{1}{\zeta \omega_n} \tag{10.7}$$

ζ and ω_n (Chapter 3) are the two most significant figures of merit, defined for second-order but often useful for higher-order systems. Specifications are often given in terms of ζ and ω_n.

This concept is developed more fully for both continuous and discrete-time systems in Chapter 14, in terms of dominant pole-zero approximations.

EXAMPLE 10.3. The plot of the unit step response of an underdamped continuous second-order system in Fig. 10-4 illustrates time-domain specifications.

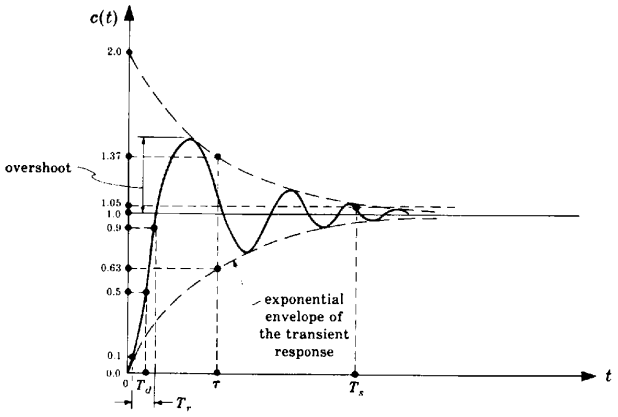


Fig. 10-4

10.5 SYSTEM COMPENSATION

We assume first that G and H are fixed configurations of components over which the designer has no control. To meet performance specifications for feedback control systems, appropriate *compensation* components (sometimes called *equalizers*) are normally introduced into the system. Compensation components may consist of either passive or active elements, several of which were discussed in Chapters 2 and 6. They may be introduced into the forward path (*cascade compensation*), or the feedback path (*feedback compensation*), as shown in Fig. 10-5:

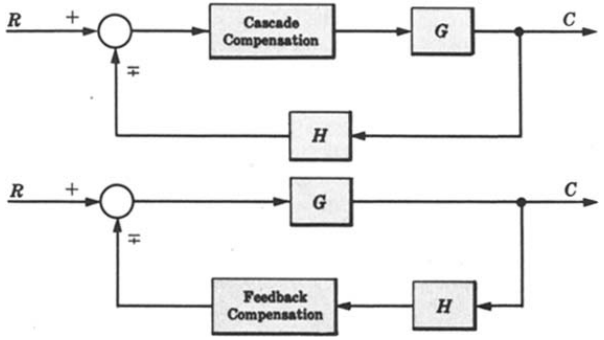


Fig. 10-5

Feedback compensation may also occur in minor feedback loops (Fig. 10-6).

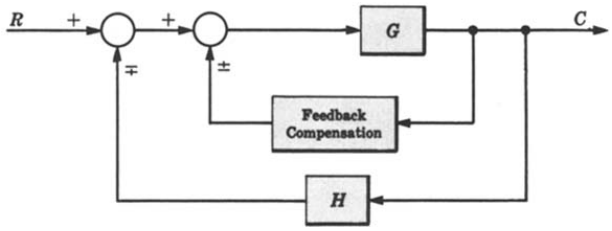


Fig. 10-6

Compensators are normally designed so that the overall system (continuous or discrete) has an acceptable transient response, and hence stability characteristics, and a desired or acceptable steady state accuracy (Chapter 9). These objectives are often conflicting, because small steady state errors usually require large open-loop gains, which typically degrade system stability. For this reason, simple compensator elements are often combined in a single design. They typically consist of combinations of components that modify the gain K and/or time constants τ, or otherwise add zeros or poles to GH. *Passive* compensators include passive physical elements such as resistive-capacitive networks, to modify K ($K < 1$), time constants, zeros, or poles; *lag*, *lead*, and *lag-lead* networks are examples (Chapter 6). The most common active compensator is the amplifier ($K > 1$). A very general one is the PID (proportional-integral-derivative) controller discussed in Chapter 2 and 6 (Examples 2.14 and 6.7), commonly used in the design of both analog (continuous) and discrete-time (digital) systems.

10.6 DESIGN METHODS

Design by analysis is the design scheme developed in this book, because it is generally a more practical approach, with the exception that direct design of digital systems, discussed in Section 10.8, is a true synthesis technique. The previously mentioned analysis methods, reiterated below, are applied to design in Chapters 12, 14, 16, and 18.

1. Nyquist Plot (Chapter 12)
2. Root-Locus (Chapter 14)
3. Bode Plot (Chapters 16)
4. Nichols Chart (Chapter 18)

Control system analysis and design procedures based on these methods have been automated in special-purpose computer software packages called *Computer-Aided Design* (CAD) packages.

Of the four methods listed above, the Nyquist, Bode, and Nichols methods are *frequency response* techniques, because in each of them the properties of $GH(\omega)$, that is, $GH(j\omega)$ for continuous systems or $GH(e^{j\omega T})$ for discrete-time systems [Equation (*10.1*)], are explored graphically as a function of angular frequency ω. More importantly, analysis and design using these methods is performed in fundamentally the same manner for continuous and discrete-time systems, as illustrated in subsequent chapters. The only differences (in specific details) stem from the fact that the stability region for continuous systems is the left half of the s-plane, and that for discrete-time systems is the unit circle in the z-plane. A transformation of variables, however, called the *w-transform*, permits analysis and design of discrete-time systems using specific results developed for continuous systems. We present the major features and the results for the w-transform in the next section, for use in analysis and design of control systems in subsequent chapters.

10.7 THE w-TRANSFORM FOR DISCRETE-TIME SYSTEMS ANALYSIS AND DESIGN USING CONTINUOUS SYSTEM METHODS

The w-transform was defined in Chapter 5 for stability analysis of discrete-time systems. It is a bilinear transformation between the complex w-plane and the complex z-plane defined by the pair:

$$w = \frac{z-1}{z+1} \qquad z = \frac{1+w}{1-w} \tag{10.8}$$

where $z = \mu + j\nu$. The complex variable w is defined as

$$w = \operatorname{Re} w + j \operatorname{Im} w \tag{10.9}$$

The following relations among these variables are useful in the analysis and design of discrete-time control systems:

1. $$\operatorname{Re} w = \frac{\mu^2 + \nu^2 - 1}{\mu^2 + \nu^2 + 2\mu + 1} \tag{10.10}$$

2. $$\operatorname{Im} w = \frac{2\nu}{\mu^2 + \nu^2 + 2\mu + 1} \tag{10.11}$$

3. $$\text{If } |z| < 1, \text{ then } \operatorname{Re} w < 0 \tag{10.12}$$

4. $$\text{If } |z| = 1, \text{ then } \operatorname{Re} w = 0 \tag{10.13}$$

5. $$\text{If } |z| > 1, \text{ then } \operatorname{Re} w > 0 \tag{10.14}$$

6. On the unit circle of the z-plane:

$$z = e^{j\omega T} = \cos \omega T + j \sin \omega T \tag{10.15}$$

$$\mu^2 + \nu^2 = \cos^2 \omega T + \sin^2 \omega T = 1 \tag{10.16}$$

$$w = j\frac{\nu}{\mu + 1} \tag{10.17}$$

Thus the region inside the unit circle in the z-plane maps into the left half of the w-plane (LHP); the region outside the unit circle maps into the right half of the w-plane (RHP); and the unit circle maps onto the imaginary axis of the w-plane. Also, rational functions of z map into rational functions of w.

For these reasons, absolute and relative stability properties of discrete systems can be determined using methods developed for continuous systems in the s-plane. Specifically, for frequency response analysis and design of discrete-time systems in the w-plane, we generally treat the w-plane as if it were the s-plane. However, we must account for distortions in certain mappings, particularly angular frequency, when interpreting the results.

From Equation (*10.17*), we define an angular frequency ω_w on the imaginary axis in the w-plane by

$$\omega_w \equiv \frac{\nu}{\mu + 1} \tag{10.18}$$

This new angular frequency ω_w in the w-plane is related to the true angular frequency ω in the z-plane by

$$\omega_w = \tan\frac{\omega T}{2} \qquad \text{or} \qquad \omega = \frac{2}{T}\tan^{-1}\omega_w \tag{10.19}$$

The following properties of ω_w are useful in plotting functions for frequency response analysis in the w-plane:

1. $$\text{If } \omega = 0, \text{ then } \omega_w = 0 \tag{10.20}$$

2. $$\text{If } \omega \to \frac{\pi}{T}, \text{ then } \omega_w \to +\infty \tag{10.21}$$

3. $$\text{If } \omega \to -\frac{\pi}{T}, \text{ then } \omega_w \to -\infty \tag{10.22}$$

4. $$\text{The range } -\frac{\pi}{T} < \omega < \frac{\pi}{T} \text{ is mapped into the range } -\infty < \omega_w < +\infty \tag{10.23}$$

Algorithm for Frequency Response Analysis and Design Using the w-Transform

The procedure is summarized as follows:

1. Substitute $(1 + w)/(1 - w)$ for z in the open-loop transfer function $GH(z)$:

$$GH(z)\big|_{z=(1+w)/(1-w)} \equiv GH'(w) \tag{10.24}$$

2. Generate frequency response curves, that is, Nyquist Plots, Bode Plots, etc., for

$$GH'(w)|_{w=j\omega_w} \equiv GH'(j\omega_w) \tag{10.25}$$

3. Analyze relative stability properties of the system in the w-plane (as if it were the s-plane). For example, determine gain and phase margins, crossover frequencies, the closed-loop frequency response, the bandwidth, or any other desired frequency-response-related characteristics.

4. Transform w-plane critical frequencies (values of ω_w) determined in Step 3 into angular frequencies (values of ω) in the true frequency domain (z-plane), using Equation (10.19).

5. If this is a design problem, design appropriate compensators to modify $GH'(j\omega_w)$ to satisfy performance specifications.

This algorithm is developed further and applied in Chapters 15 through 18.

EXAMPLE 10.4. The open-loop transfer function

$$GH(z) = \frac{(z+1)^2/100}{(z-1)(z+\frac{1}{3})(z+\frac{1}{2})} \tag{10.26}$$

is transformed into the w-domain by substituting $z = (1+w)/(1-w)$ in the expression for $GH(z)$, which yields

$$GH'(w) = \frac{-6(w-1)/100}{w(w+2)(w+3)} \tag{10.27}$$

Relative stability analysis of $GH'(w)$ is postponed until Chapter 15.

10.8 ALGEBRAIC DESIGN OF DIGITAL SYSTEMS, INCLUDING DEADBEAT SYSTEMS

When digital computers or microprocessors are components of a discrete-time system, compensators can be readily implemented in software or firmware, thereby facilitating direct design of the system by algebraic solution for the transfer function of the compensator that satisfies given design objectives. For example, suppose we wish to construct a system having a given closed-loop transfer function C/R, which might be defined by requisite closed-loop characteristics such as bandwidth, steady state gain, response time, etc. Then, given the plant transfer function $G_2(z)$, the required forward loop compensator $G_1(z)$ can be determined from the relation for the closed-loop transfer function of the canonical system given in Section 7.5:

$$\frac{C}{R} = \frac{G_1G_2}{1+G_1G_2H} \tag{10.28}$$

Then the required compensator is determined by solving for $G_1(z)$:

$$G_1 = \frac{C/R}{G_2(1-HC/R)} \tag{10.29}$$

EXAMPLE 10.5. The unity feedback ($H \equiv 1$) system in Fig. 10-7, with $T = 0.1$-sec uniform and synchronous sampling, is required to have a steady state gain $(C/R)(1) = 1$ and a rise time T_r of 2 sec or less.

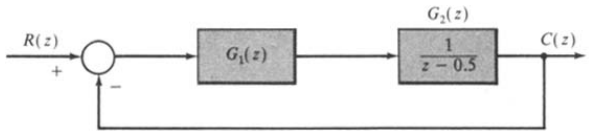


Fig. 10-7

The simplest C/R that satisfies the requirements is $(C/R) = 1$. However, the required compensator would be

$$G_1 = \frac{\frac{C}{R}}{G_2\left(1-\frac{C}{R}\right)} = \frac{1}{\frac{1}{z-0.5}(1-1)} = \frac{z-0.5}{0}$$

which has infinite gain, a zero at $z = 0.5$, and no poles, which is unrealizable. For realizability (Section 6.6), G_1 must have at least as many poles as zeros. Consequently, even with cancellation of the poles and zeros of G_2 by zeros and poles of G_1, C/R must contain at least $n-m$ poles, where n is the number of poles and m is the number of zeros of G_2.

The simplest realizable C/R has the form:

$$\frac{C}{R} = \frac{K}{z-a}$$

As shown in Problem 10.10, the rise time for a first-order discrete-time system, like the one given by C/R above, is

$$T_r \le \frac{T \ln\frac{1}{9}}{\ln a}$$

Solving for a, we get

$$a = \left[\frac{1}{9}\right]^{T_r/T} = \left[\frac{1}{9}\right]^{20} = 0.8959$$

Then

$$\frac{C}{R} = \frac{K}{z-a} = \frac{K}{z-0.8959}$$

and, for the steady state gain $(C/R)(1)$ to be 1, $K = 1 - 0.8959 = 0.1041$. Therefore the required compensator is

$$G_1 = \frac{\frac{C}{R}}{G_2\left(1-\frac{C}{R}\right)} = \frac{\frac{0.1041}{z-0.8959}}{\frac{1}{z-0.5}\left(1-\frac{0.1041}{z-0.8959}\right)} = \frac{0.1041(z-0.5)}{z-1}$$

We see that G_1 has added a pole to G_1G_2 at $z = 1$, making the system type 1. This is due to the requirement that the steady state gain equal 1.

Deadbeat systems are a class of discrete-time systems that can be readily designed using the direct approach described above. By definition, the closed-loop *transient* response of a **deadbeat system** has finite length, that is, it becomes zero, and remains zero, after a finite number of sample times. In response to a step input, the output of such a system is constant at each sample time after a finite period. This is termed a **deadbeat response**.

EXAMPLE 10.6. For a unity feedback system with forward transfer function

$$G_2(z) = \frac{K_1(z+z_1)}{(z+p_1)(z+p_2)}$$

introduction of a feedforward compensator with

$$G_1(z) = \frac{(z+p_1)(z+p_2)}{(z-K_1)(z+z_1)}$$

results in the closed-loop transfer function:

$$\frac{C}{R} = \frac{G_1G_2}{1+G_1G_2} = \frac{K_1}{z}$$

The impulse response of this system is $c(0) = K_1$ and $c(k) = 0$ for $k > 0$. The step response is $c(0) = 0$ and $c(k) = K_1$ for $k > 0$.

In general, systems can be designed to exhibit a deadbeat response with a transient response $n - m$ samples long, where m is the number of zeros and n is the number of poles of the plant. However, to avoid *intersample ripple* (periodic or aperiodic variations) in mixed continuous/discrete-time systems, where $G_2(z)$ has a continuous input and/or output, the zeros of $G_2(z)$ should not be cancelled by the compensator as in Example 10.5. The transient response in these cases is a minimum of n samples in length and the closed loop transfer function has n poles at $z = 0$.

EXAMPLE 10.7. For a system with

$$G_2(z) = \frac{K(z+0.5)}{(z-0.2)(z-0.4)}$$

let

$$G_1(z) = \frac{(z-0.2)(z-0.4)}{(z+a)(z+b)}$$

Then

$$\frac{C}{R} = \frac{G_1G_2}{1+G_1G_2} = \frac{K(z+0.5)}{(z+a)(z+b)+K(z+0.5)}$$
$$= \frac{K(z+0.5)}{z^2+(a+b+K)z+ab+0.5K}$$

For a deadbeat response, we choose

$$\frac{C}{R} \equiv \frac{K(z+0.5)}{z^2}$$

and therefore

$$a+b+K \equiv 0$$
$$ab+0.5K \equiv 0$$

There are many possible solutions for a, b, and K and one is $a = 0.3$, $b = -0.75$, and $K = 0.45$.

If it is required that the closed-loop system be type l, it is necessary that $G_1(z)G_2(z)$ contain l poles at $z = 1$. If $G_2(z)$ has the required number of poles, they should be retained, that is, not cancelled by zeros of $G_1(z)$. If $G_2(z)$ does not have all the required poles at $z = 1$, they can be added in $G_1(z)$.

EXAMPLE 10.8. For the system with

$$G_2(z) = \frac{K}{z-1}$$

suppose a type 2 closed-loop system with deadbeat response is desired. This can be achieved with a compensator of the form:

$$G_1(z) = \frac{z+a}{z-1}$$

which adds a pole at $z = 1$. Then

$$\frac{C}{R} = \frac{G_1G_2}{1+G_1G_2} = \frac{K(z+a)}{(z-1)^2+K(z+a)} = \frac{K(z+a)}{z^2+(K-2)z+1+Ka}$$

If a deadbeat response is desired, we must have

$$\frac{C}{R} = \frac{K(z+a)}{z^2}$$

and therefore $K - 2 = 0$ and $1 + Ka = 0$, giving $K = 2$ and $a = -0.5$.

Solved Problems

10.1. The graph of Fig. 10-8 represents the input-output characteristic of a controller-amplifier for a feedback control system whose other components are linear. What is the linear range of $e(t)$ for this system?

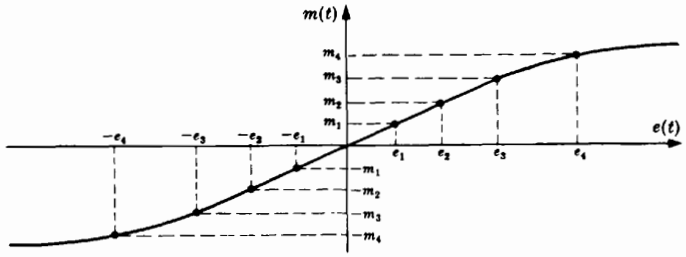


Fig. 10-8

The amplifier-controller operates linearly over the approximate range $-e_3 \le e \le e_3$.

10.2. Determine the gain margin for the system in which $GH(j\omega) = 1/(j\omega+1)^3$.

Writing $GH(j\omega)$ in polar form, we have

$$GH(j\omega) = \frac{1}{(\omega^2+1)^{3/2}} \angle -3\tan^{-1}\omega \qquad \arg GH(j\omega) = -3\tan^{-1}\omega$$

Then $-3\tan^{-1}\omega_\pi = -\pi$, $\omega_\pi = \tan(\pi/3) = 1.732$. Hence, by Equation (*10.2*), gain margin $= 1/|GH(j\omega_\pi)| = 8$.

10.3. Determine the phase margin for the system of Problem 10.2.

We have

$$|GH(j\omega)| = \frac{1}{(\omega^2+1)^{3/2}} = 1$$

only when $\omega = \omega_1 = 0$. Therefore

$$\phi_{PM} = 180° + (-3\tan^{-1}0) = 180° = \pi \text{ radians}$$

10.4. Determine the average value of $T_d(\omega)$ over the frequency range $0 \le \omega \le 10$ for $C/R = j\omega/(j\omega+1)$. $T_d(\omega)$ is given by Equation (*10.4*).

$$\gamma = \arg\frac{C}{R}(j\omega) = \frac{\pi}{2} - \tan^{-1}\omega \qquad \text{and} \qquad T_d(\omega) = \frac{-d\gamma}{d\omega} = \frac{d}{d\omega}[\tan^{-1}\omega] = \frac{1}{1+\omega^2}$$

Therefore

$$\text{Avg } T_d(\omega) = \frac{1}{10}\int_0^{10}\frac{d\omega}{1+\omega^2} = 0.147 \text{ sec}$$

10.5. Determine the bandwidth for the system with transfer function $(C/R)(s) = 1/(s+1)$.

We have

$$\left|\frac{C}{R}(j\omega)\right| = \frac{1}{\sqrt{\omega^2+1}}$$

A sketch of $|(C/R)(j\omega)|$ versus ω is given in Fig. 10-9.

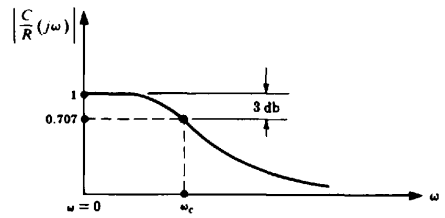


Fig. 10-9

ω_c is determined from $1/\sqrt{\omega_c^2 + 1} = 0.707$. Since $|(C/R)(j\omega)|$ is a strictly decreasing function of positive frequency, we have BW $= \omega_c = 1$ rad.

10.6. How many octaves are between (*a*) 200 Hz and 800 Hz, (*b*) 200 Hz and 100 Hz, (*c*) 10,048 rad/sec (rps) and 100 Hz?

(*a*) Two octaves.

(*b*) One octave.

(*c*) $f = \omega/2\pi = 10{,}048/2\pi = 1600$ Hz. Hence there are four octaves between 10,048 rps and 100 Hz.

10.7. Determine the resonance peak M_p and the resonant frequency ω_p for the system whose transfer function is $(C/R)(s) = 5/(s^2 + 2s + 5)$.

$$\left|\frac{C}{R}(j\omega)\right| = \frac{5}{|-\omega^2 + 2j\omega + 5|} = \frac{5}{\sqrt{\omega^4 - 6\omega^2 + 25}}$$

Setting the derivative of $|(C/R)(j\omega|$ equal to zero, we get $\omega_p = \pm\sqrt{3}$. Therefore

$$M_p = \max_{\omega}\left|\frac{C}{R}(j\omega)\right| = \left|\frac{C}{R}(j\sqrt{3})\right| = \frac{5}{4}$$

10.8. The output in response to a unit step function input for a particular continuous control system is $c(t) = 1 - e^{-t}$. What is the delay time T_d?

The output is given as a function of time. Therefore, the time-domain definition of T_d presented in Section 10.4 is applicable. The final value of the output is $\lim_{t\to\infty} c(t) = 1$. Hence T_d (at 50% of the final value) is the solution of $0.5 = 1 - e^{-T_d}$, and is equal to $\log_e(2)$, or 0.693.

10.9. Find the rise time T_r for $c(t) = 1 - e^{-t}$.

At 10% of the final value, $0.1 = 1 - e^{-t_1}$; hence $t_1 = 0.104$ sec. At 90% of the final value, $0.9 = 1 - e^{-t_2}$; thus $t_2 = 2.302$ sec. Then $T_r = 2.302 - 0.104 = 2.198$ sec.

10.10. Determine the rise time of the first-order discrete system

$$P(z) = (1 - a)/(z - a) \text{ with } |a| < 1.$$

For a step input, the output transform is

$$Y(z) = P(z)U(z) = \frac{(1-a)z}{(z-1)(z-a)}$$

and the time response is $y(k) = 1 - a^k$ for $k = 0, 1, \ldots$. Since $y(\infty) = 1$, the rise time T_r is the time required for this unit step response to go from 0.1 to 0.9. Since the sampled response may not have the exact values 0.1 and 0.9, we must find the sampled values that bound these values. Thus, for the lower value, $y(k) \le 0.1$, or $1 - a^k \le 0.1$ and therefore $a^k \ge 0.9$. Similarly for $y(k + T_r/T) = 1 - a^{k+T_r/T} \ge 0.9$, $a^{k+T_r/T} \le 0.1$.

Dividing the two expressions, we get

$$\frac{a^{k+T_r/T}}{a^k} \le \frac{1}{9}$$

or

$$a^{T_r/T} \le \frac{1}{9}$$

Then, by taking logarithms of both sides, we get

$$T_r \le \frac{T \ln\frac{1}{9}}{\ln a}$$

10.11. Verify the six properties of the w-transform in Section 10.7, Equations (*10.10*) through (*10.17*).

From $w = (z - 1)/(z + 1)$ and $z = \mu + j\nu$,

$$w = \frac{\mu + j\nu - 1}{\mu + j\nu + 1} = \frac{(\mu - 1 + j\nu)(\mu + 1 - j\nu)}{(\mu + 1 + j\nu)(\mu + 1 - j\nu)} = \left(\frac{\mu^2 + \nu^2 - 1}{\mu^2 + \nu^2 + 2\mu + 1}\right) + j\left(\frac{2\nu}{\mu^2 + \nu^2 + 2\mu + 1}\right)$$

Thus

1. $$\text{Re } w = \frac{\mu^2 + \nu^2 - 1}{\mu^2 + \nu^2 + 2\mu + 1} \equiv \sigma_w$$
2. $$\text{Im } w = \frac{2\nu}{\mu^2 + \nu^2 + 2\mu + 1} \equiv \omega_w$$
3. $|z| < 1$ means $\mu^2 + \nu^2 < 1$, which implies $\sigma_w < 0$
4. $|z| = 1$ means $\mu^2 + \nu^2 = 1$, which implies $\sigma_w = 0$
5. $|z| > 1$ means $\mu^2 + \nu^2 > 1$, which implies $\sigma_w > 0$

The sixth property follows from elementary trigonometric identities.

10.12. Show that the transformed angular frequency ω_w is related to the real frequency ω by Equation (*10.19*).

From Problem 10.11, $|z| = 1$ also implies that $w = j[\nu/(\mu + 1)] \equiv j\omega_w$ [Equation (*10.17*)]. But $|z| = 1$ implies that $z = e^{j\omega T} \equiv \cos\omega T + j\sin\omega T = \mu + j\nu$ [Equation (*10.15*)]. Therefore

$$\omega_w = \frac{\sin\omega T}{\cos\omega T + 1}$$

Finally, substituting the following half-angle identities of trigonometry into the last expression:

$$2\sin\left(\frac{\omega T}{2}\right)\cos\left(\frac{\omega T}{2}\right) = \sin\omega T$$

$$\cos^2\left(\frac{\omega T}{2}\right) - \sin^2\left(\frac{\omega T}{2}\right) = \cos\omega T$$

$$\cos^2\left(\frac{\omega T}{2}\right) + \sin^2\left(\frac{\omega T}{2}\right) = 1$$

we have

$$\omega_w = \frac{2\sin\left(\frac{\omega T}{2}\right)\cos\left(\frac{\omega T}{2}\right)}{2\cos^2\left(\frac{\omega T}{2}\right)} = \frac{\sin\left(\frac{\omega T}{2}\right)}{\cos\left(\frac{\omega T}{2}\right)} = \tan\left(\frac{\omega T}{2}\right)$$

10.13. For the uniformly and synchronously sampled system given in Fig. 10-10, determine $G_1(z)$ so that the system is type 1 with a deadbeat response.

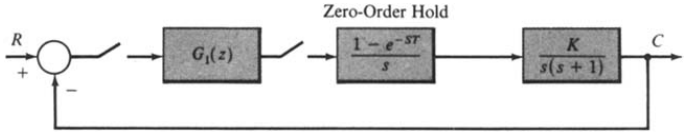


Fig. 10-10

The forward loop z-transform, assuming *fictitious* sampling of the output $c(t)$ (see Section 6.8), is determined from Equation (6.9):

$$G_2(z) = \frac{z-1}{z} \mathcal{Z}\left\{\mathcal{L}^{-1}\left(\frac{G(s)}{s}\right)\Big|_{t=kT}\right\} = \frac{K_1(z+z_1)}{(z-1)(z-e^{-T})}$$

where

$$K_1 \equiv K(T+e^{-T}-1) \qquad \text{and} \qquad z_1 \equiv \frac{1-e^{-T}-Te^{-T}}{T+e^{-T}-1}$$

Let $G_1(z)$ have the form $G_1(z) = (z-e^{-T})/(z+b)$. Then, if we also assume a fictitious sampler at the input $r(t)$, we can determine the closed-loop z-domain transfer function:

$$\frac{C}{R} = \frac{G_1G_2}{1+G_1G_2} = \frac{K_1(z+z_1)}{(z-1)(z+b)+K_1(z+z_1)}$$

$$= \frac{K_1(z+z_1)}{z^2+(b-1+K_1)z-b+K_1z_1}$$

For a deadbeat response, $b-1+K_1=0$ $(b=1-K_1)$ and $-b+K_1z_1=0$ $(-1+K_1+K_1z_1=0)$. Then

$$K_1 = \frac{1}{1+z_1}$$

and

$$b = 1-K_1 = \frac{z_1}{1+z_1}$$

Since $K_1 = K(T+e^{-T}-1)$,

$$K = \frac{K_1}{T+e^{-T}-1} = \frac{1}{(1+z_1)(T+e^{-T}-1)} = \frac{1}{T(1-e^{-T})}$$

For this system, with continuous input and output signals, $(C/R)(z)$ determined above gives the closed-loop input-output relationship at the sampling times only.

Supplementary Problems

10.14. Determine the phase margin for $GH = 2(s+1)/s^2$.

10.15. Find the bandwidth for $GH = 60/s(s+2)(s+6)$ for the closed-loop system.

10.16. Calculate the gain and phase margin for $GH = 432/s(s^2+13s+115)$.

10.17. Calculate the phase margin and bandwidth for $GH = 640/s(s+4)(s+16)$ for the closed-loop system.

Answers to Supplementary Problems

10.14. $\phi_{PM} = 65.5°$

10.15. BW = 3 rad/sec

10.16. Gain margin = 3.4, phase margin = 65°

10.17. $\phi_{PM} = 17°$, BW = 5.5 rad/sec

Chapter 11

Nyquist Analysis

11.1 INTRODUCTION

Nyquist analysis, a frequency response method, is essentially a graphical procedure for determining absolute and relative stability of closed-loop control systems. Information about stability is available directly from a graph of the open-loop frequency response function $GH(\omega)$, once the feedback system has been put into canonical form.

Nyquist methods are applicable to both continuous and discrete-time control systems, and the methodological development for Nyquist analysis is presented here for both types of systems, with some emphasis given to continuous systems, for pedagogical purposes.

There are several reasons why the Nyquist method may be chosen to determine information about system stability. The methods of Chapter 5 (Routh, Hurwitz, etc.) are often inadequate because, with few exceptions, they can only be used for determining *absolute* stability, and are only applicable to systems whose characteristic equation is a *finite polynomial* in s or z. For example, when a signal is delayed by T seconds somewhere in the loop of a continuous system, exponential terms of the form e^{-Ts} appear in the characteristic equation. The methods of Chapter 5 can be applied to such systems if e^{-Ts} is approximated by a few terms of the power series

$$e^{-Ts} = 1 - Ts + \frac{T^2s^2}{2!} - \frac{T^3s^3}{3!} + \cdots$$

but this technique yields only *approximate* stability information. The Nyquist method handles systems with time delays without the necessity of approximations, and hence yields *exact* results about both absolute and relative stability of the system.

Nyquist techniques are also useful for obtaining information about transfer functions of components or systems from experimental frequency response data. The Polar Plot (Section 11.5) may be directly graphed from sinusoidal steady state measurements on the components making up the open-loop transfer function. This feature is very useful in the determination of system stability characteristics when transfer functions of loop components are not available in analytic form, or when physical systems are to be tested and evaluated experimentally.

In the next several sections we present the mathematical preliminaries and techniques necessary for generating Polar Plots and Nyquist Stability Plots of feedback control systems, and the mathematical basis and properties of the Nyquist Stability Criterion. The remaining sections of this chapter deal with the interpretation and uses of Nyquist analysis for the determination of *relative* stability and evaluation of the closed-loop frequency response.

11.2 PLOTTING COMPLEX FUNCTIONS OF A COMPLEX VARIABLE

A real function of a real variable is easily graphed on a single set of coordinate axes. For example, the real function $f(x)$, x real, is easily plotted in rectangular coordinates with x as the abscissa and $f(x)$ as the ordinate. A complex function of a complex variable, such as the transfer function $P(s)$ with $s = \sigma + j\omega$, cannot be plotted on a single set of coordinates.

The complex variable $s = \sigma + j\omega$ depends on two independent quantities, the real and imaginary parts of s. Hence s cannot be represented by a line. The complex function $P(s)$ also has real and imaginary parts. It too cannot be graphed in a single dimension. Similarly, the complex variable $z = \mu + j\nu$ and discrete-time system complex transfer functions $P(z)$ cannot be graphed in one dimension.

In general, in order to plot $P(s)$ with $s = \sigma + j\omega$, two two-dimensional graphs are required. The first is a graph of $j\omega$ versus σ called the **s-plane**, the same set of coordinates as those used for plotting pole-zero maps in Chapter 4. The second is the imaginary part of $P(s)$ (Im P) versus the real part of $P(s)$ (Re P) called the **$P(s)$-plane**. The corresponding coordinate planes for discrete-time systems are the **z-plane** and the **$P(z)$-plane**.

The correspondence between points in the two planes is called a **mapping** or **transformation**. For example, points in the s-plane are *mapped* into points of the $P(s)$-plane by the function P (Fig. 11-1).

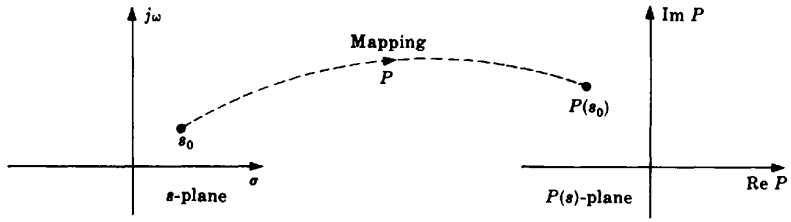


Fig. 11-1

In general, only a very specific locus of points in the s-plane (or the z-plane) is mapped into the $P(s)$-plane [or the $P(z)$-plane]. For Nyquist Stability Plots this locus is called the *Nyquist Path*, the subject of Section 11.7.

For the special case $\sigma = 0$, $s = j\omega$, the s-plane degenerates into a line, and $P(j\omega)$ may be represented in a $P(j\omega)$-plane with ω as a parameter. *Polar Plots* are constructed in the $P(j\omega)$-plane from this line ($s = j\omega$) in the s-plane.

EXAMPLE 11.1. Consider the complex function $P(s) = s^2 + 1$. The point $s_0 = 2 + j4$ is mapped into the point $P(s_0) = P(2 + j4) = (2 + j4)^2 + 1 = -11 + j16$ (Fig. 11-2).

Fig. 11-2

11.3 DEFINITIONS

The following definitions are required in subsequent sections.

Definition 11.1: If the *derivative* of P at s_0 defined by

$$\left.\frac{dP}{ds}\right|_{s=s_0} \equiv \lim_{s \to s_0} \left[\frac{P(s) - P(s_0)}{s - s_0}\right]$$

exists at all points in a region of the s-plane, that is, if the limit is finite and unique, then P is **analytic** in that region [same definition for $P(z)$ in the z-plane, with z replacing s and z_0 replacing s_0].

Transfer functions of practical physical systems (those considered in this book) are analytic in the finite s-plane (or finite z-plane) except at the poles of $P(s)$ [or poles of $P(z)$]. In subsequent developments, when there is no danger of ambiguity, and when a given statement applies to both $P(s)$ and $P(z)$, then $P(s)$ or $P(z)$ may be abbreviated as P with no argument.

Definition 11.2: A point at which P [$P(s)$ or $P(z)$] is not analytic is a **singular point** or **singularity** of P [$P(s)$ or $P(z)$].

A *pole* of P [$P(s)$ or $P(z)$] is a singular point.

Definition 11.3: A **closed contour** in a complex plane is a continuous curve beginning and ending at the same point (Fig. 11-3).

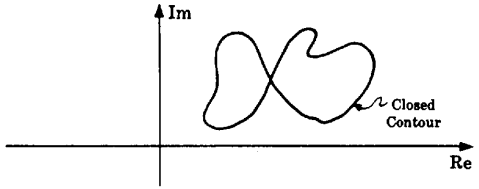


Fig. 11-3

Definition 11.4: All points to the right of a contour as it is traversed in a prescribed direction are said to be **enclosed** by it (Fig. 11-4).

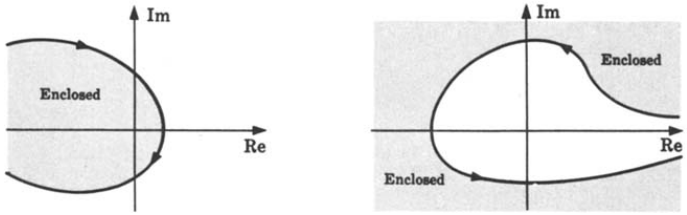


Fig. 11-4

Definition 11.5: A *clockwise* (CW) traverse around a contour is defined as the **positive direction** (Fig. 11-5).

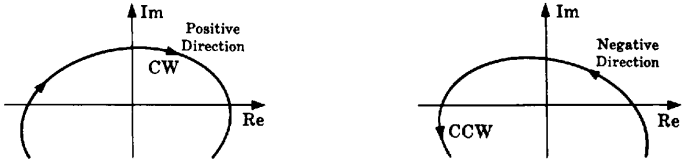


Fig. 11-5

Definition 11.6: A closed contour in the P-plane is said to make n **positive encirclements** of the origin if a radial line drawn from the origin to a point on the P curve rotates in a clockwise (CW) direction through $360n$ degrees in completely traversing the closed path. If the path is traversed in a counterclockwise (CCW) direction, a **negative encirclement** is obtained. The **total number of encirclements** N_0 is equal to the CW minus the CCW encirclements.

EXAMPLE 11.2. The P-plane contour in Fig. 11-6 encircles the origin once. That is, $N_0 = 1$. Beginning at point a, we rotate a radial line from the origin to the contour in a CW direction to point c. The angle subtended is $+270°$. From c to d the angle increases, then decreases, and the sum total is $0°$. From d to e and back to d again, the angle swept out by the radial line is again $0°$. d to c is $0°$ and c to a is clearly $+90°$. Hence the total angle is $270° + 90° = 360°$. Therefore $N_0 = 1$.

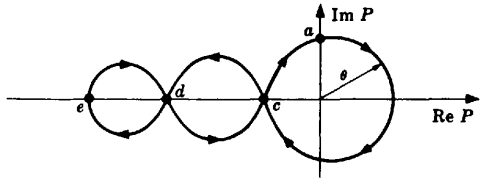


Fig. 11-6

11.4 PROPERTIES OF THE MAPPING $P(s)$ or $P(z)$

All mappings P [$P(s)$ or $P(z)$] considered in the remainder of this chapter have the following properties.

1. P is a *single-valued function*. That is, every point in the s-plane (or the z-plane) maps into one and only one point in the P-plane.
2. s-plane (z-plane) contours avoid singular points of P.
3. P is *analytic* except possibly at a finite number of points (singularities) in the s-plane (or the z-plane).
4. Every closed contour in the s-plane (or the z-plane) maps into a closed contour in the P-plane.
5. P is a *conformal mapping*. This means that the direction of and the angle between any two intersecting curves at their point of intersection in the s-plane (or the z-plane) are preserved by the mapping of these curves into the P-plane.
6. The mapping P obeys the *principle of arguments*. That is, the *total number of encirclements* N_0 *of the origin* made by a closed P contour in the P-plane, mapped from a closed s-plane (or z-plane) contour, is equal to the number of zeros Z_0 minus the number of poles P_0 of P enclosed by the the s-plane (or z-plane) contour. That is,

$$N_0 = Z_0 - P_0 \qquad (11.1)$$

7. If the origin is *enclosed* by the P contour, then $N_0 > 0$. If the origin is *not enclosed* by the P contour, then $N_0 \leq 0$. That is,

$$\text{enclosed} \Rightarrow N_0 > 0$$
$$\text{not enclosed} \Rightarrow N_0 \leq 0$$

The *sign of* N_0 is easily determined by shading the region to the right of the contour in the prescribed direction. If the origin falls in a shaded region, $N_0 > 0$; if not, $N_0 \leq 0$.

EXAMPLE 11.3. The principle of conformal mapping is illustrated in Fig. 11-7. Curves C_1 and C_2 are mapped into C_1' and C_2'. The angle between the tangents to these curves at s_0 and $P(s_0)$ is equal to α, and the curves turn right at s_0 and at $P(s_0)$, as indicated by the arrows in both graphs.

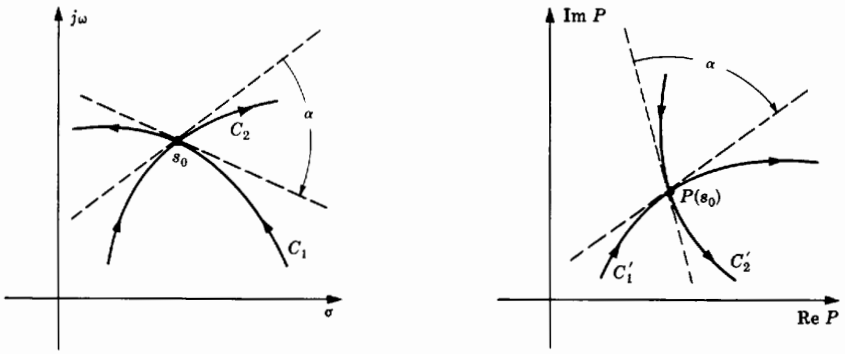


Fig. 11-7

EXAMPLE 11.4. A certain transfer function $P(s)$ is known to have one zero in the right half of the s-plane, and this zero is enclosed by the s-plane contour mapped into the $P(s)$-plane in Fig. 11-8. Points s_1, s_2, s_3 and $P(s_1), P(s_2), P(s_3)$ determine the directions of their respective contours. The shaded region to the right of the $P(s)$-plane contour indicates that $N_0 \leq 0$, since the origin does not lie in the shaded region. But, clearly, the $P(s)$ contour encircles the origin once in a CCW direction. Hence $N_0 = -1$. Thus the number of poles of $P(s)$ enclosed by the s-plane contour is $P_0 = Z_0 - N_0 = 1 - (-1) = 2$.

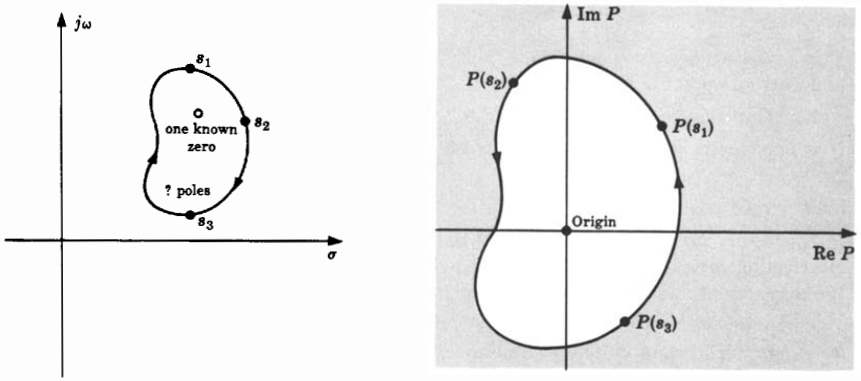


Fig. 11-8

11.5 POLAR PLOTS

A continuous system transfer function $P(s)$ may be represented in the frequency domain as a sinusoidal transfer function by substituting $j\omega$ for s in the expression for $P(s)$. The resulting form $P(j\omega)$ is a complex function of the single variable ω. Therefore it may be plotted in two dimensions, with ω as a parameter, and written in the following equivalent forms:

$$\textbf{Polar Form:} \qquad P(j\omega) = |P(j\omega)| \angle \phi(\omega) \tag{11.2}$$

$$\textbf{Euler Form:} \qquad P(j\omega) = |P(j\omega)|(\cos\phi(\omega) + j\sin\phi(\omega)) \tag{11.3}$$

$|P(j\omega)|$ is the **magnitude** of the complex function $P(j\omega)$, and $\phi(j\omega)$ is its **phase angle**, $\arg P(j\omega)$.

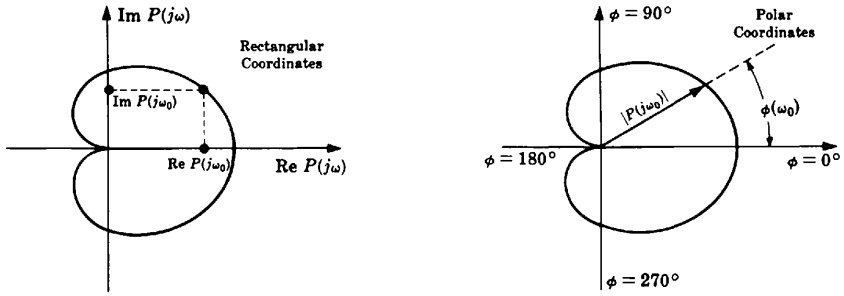
$|P(j\omega)|\cos\phi(\omega)$ is the *real part*, and $|P(j\omega)|\sin\phi(\omega)$ is the *imaginary part* of $P(j\omega)$. Therefore $P(j\omega)$ may also be written as

$$\textbf{Rectangular or Complex Form:} \qquad P(j\omega) = \operatorname{Re} P(j\omega) + j \operatorname{Im} P(j\omega) \tag{11.4}$$

A **Polar Plot** of $P(j\omega)$ is a graph of $\operatorname{Im} P(j\omega)$ versus $\operatorname{Re} P(j\omega)$ in the finite portion of the $P(j\omega)$-plane for $-\infty < \omega < \infty$. At singular points of $P(j\omega)$ (poles on the $j\omega$-axis), $|P(j\omega)| \to \infty$. A Polar Plot may also be generated on polar coordinate paper. The magnitude and phase angle of $P(j\omega)$ are plotted with ω varying from $-\infty$ to $+\infty$.

The locus of $P(j\omega)$ is identical on either rectangular or polar coordinates. The choice of coordinate system may depend on whether $P(j\omega)$ is available in analytic form or as experimental data. If $P(j\omega)$ is expressed analytically, the choice of coordinates depends on whether it is easier to write $P(j\omega)$ in the form of Equation (*11.2*), in which case polar coordinates are used, or in the form of Equation (*11.4*) for rectangular coordinates. Experimental data on $P(j\omega)$ are usually expressed in terms of magnitude and phase angle. In this case, polar coordinates are the natural choice.

EXAMPLE 11.5. The Polar Plots in Fig. 11-9 are identical; only the coordinate systems are different.



$P(j\omega) = \operatorname{Re} P(j\omega) + j \operatorname{Im} P(j\omega)$ $\qquad$ $P(j\omega) = |P(j\omega)| \angle \phi(\omega)$

Fig. 11-9

For *discrete-time systems*, Polar Plots are defined in the frequency domain in the same manner. Recall that we can write $z \equiv e^{sT}$ (see Section 4.9). Therefore a discrete transfer function $P(z) \equiv P(e^{sT})$ and, if we set $s = j\omega$, $P(z)$ becomes $P(e^{j\omega T})$. The **Polar Plot of $P(e^{j\omega T})$** is a graph of $\operatorname{Im} P(e^{j\omega T})$ versus $\operatorname{Re} P(e^{j\omega T})$ in the finite portion of the $P(e^{j\omega T})$-plane, for $-\infty < \omega < \infty$.

We often discuss Polar Plots, their properties, and many results dependent on these in subsequent sections in a unified manner for both continuous and discrete-time systems. To do this, we adopt for our general transfer function P the unified representation for frequency response functions given in Equation (*10.1*) for GH, that is, we use the generic representation $P(\omega)$ defined by

$$P(\omega) = \begin{cases} P(j\omega) & \text{for continuous systems} \\ P(e^{j\omega T}) & \text{for discrete-time systems} \end{cases}$$

In these terms, Equations (*11.2*) through (*11.4*) become

$$P(\omega) = |P(\omega)| \angle \phi(\omega) = |P(\omega)|(\cos\phi(\omega) + j\sin\phi(\omega)) = \operatorname{Re} P(\omega) + j \operatorname{Im} P(\omega)$$

We use this unified notation in much of the remainder of this chapter, and in subsequent chapters, particularly where the results are applicable to both continuous and discrete-time systems.

11.6 PROPERTIES OF POLAR PLOTS

The following are several useful properties of Polar Plots of $P(\omega)$ [$P(j\omega)$ or $P(e^{j\omega T})$].

1. The Polar Plot for
$$P(\omega) + a$$
where a is any complex constant, is identical to the plot for $P(\omega)$ with the origin of coordinates shifted to the point $-a = -(\text{Re } a + j \text{ Im } a)$.
2. The Polar Plot of the transfer function of a time-invariant, linear system exhibits *conjugate symmetry*. That is, the graph for $-\infty < \omega < 0$ is the mirror image about the horizontal axis of the graph for $0 \le \omega < \infty$.
3. The Polar Plot may be constructed directly from a Bode Plot (Chapter 15), if one is available. Values of magnitude and phase angle at various frequencies ω on the Bode Plot represent points along the locus of the Polar Plot.
4. Constant increments of frequency are not generally separated by equal intervals along the Polar Plot.

EXAMPLE 11.6. For $a = 1$ and $P = GH$, the Polar Plot of the function $1 + GH$ is given by the plot for GH, with the origin of coordinates shifted to the point $-1 + j0$ in rectangular coordinates (Fig. 11-10).

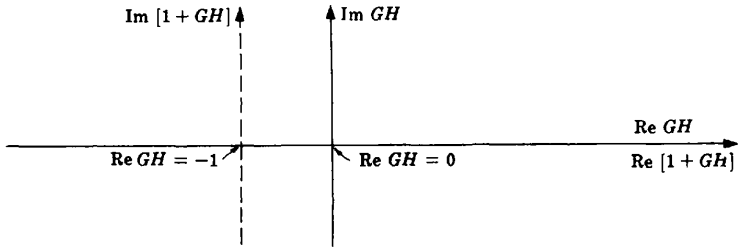


Fig. 11-10

EXAMPLE 11.7. To illustrate plotting of a transfer function, consider the open-loop continuous system transfer function

$$GH(s) = \frac{1}{s+1}$$

Letting $s = j\omega$ and rewriting $GH(j\omega)$ in the form of Equation (*11.2*) (polar form), we have

$$GH(j\omega) = \frac{1}{j\omega + 1} = \frac{1}{\sqrt{\omega^2 + 1}} \angle -\tan^{-1}\omega$$

For $\omega = 0$, $\omega = 1$, and $\omega \to \infty$:

$$GH(j0) = 1 \angle 0°$$

$$GH(j1) = (1/\sqrt{2}) \angle -45°$$

$$\lim_{\omega \to \infty} GH(j\omega) = 0 \angle -90°$$

Substitution of several other positive values of ω yields a semicircular locus for $0 \le \omega < \infty$. The graph for $-\infty < \omega < 0$ is the mirror image about the diameter of this semicircle. It is shown in Fig. 11-11 by a dashed line. Note the strikingly unequal increments of frequency between the arcs $\overline{ab}$ and $\overline{bc}$.

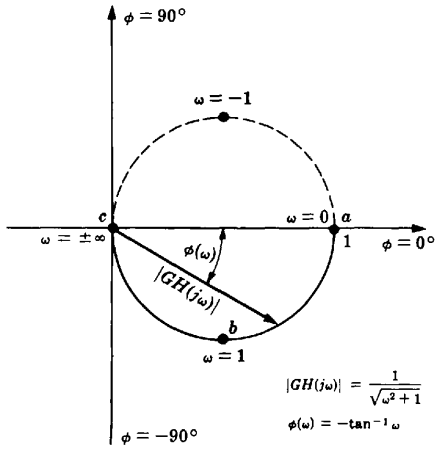


Fig. 11-11

Polar Plots are not very difficult to sketch for very simple transfer functions, although they are usually a little more difficult to determine for discrete-time systems, as illustrated in Example 11.11. But the computations can be very laborious for complicated $P(s)$ or $P(z)$. On the other hand, widely available computer programs for frequency response analysis, or more generally for plotting complex functions of a complex variable, typically generate accurate Polar Plots quite conveniently.

11.7 THE NYQUIST PATH

For continuous systems, the **Nyquist Path** is a closed contour in the s-plane, enclosing the entire right half of the s-plane (RHP). For discrete-time systems, the corresponding **Nyquist Path** encloses the entire z-plane *outside* the unit circle.

For continuous systems, in order that the Nyquist Path should not pass through any poles of $P(s)$, small semicircles along the imaginary axis or at the origin of $P(s)$ are required in the path if $P(s)$ has poles on the $j\omega$-axis or at the origin. The radii ρ of these small circles are interpreted as approaching zero in the limit.

To enclose the RHP at infinity, and thus any poles in the interior of the RHP, a large semicircular path is drawn in the RHP and the radius R of this semicircle is interpreted as being infinite in the limit.

The **generalized Nyquist Path in the *s*-plane** is illustrated by the *s*-plane contour in Fig. 11-12. It is apparent that *every pole and zero of $P(s)$ in the RHP is enclosed by the Nyquist Path* when it is mapped into the $P(s)$-plane.

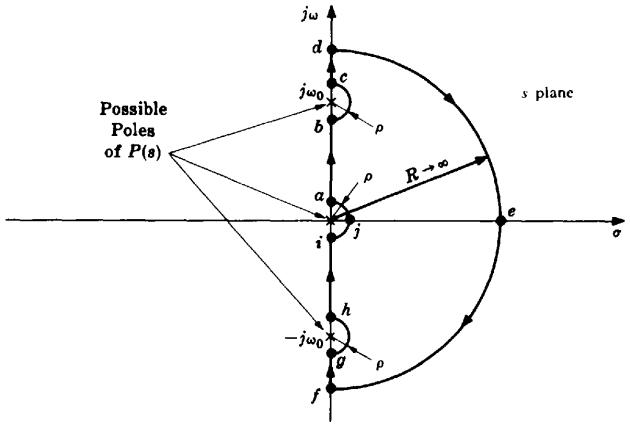


Fig. 11-12

The various portions of the Nyquist Path can be described analytically in the following manner.

Path $\overline{ab}$:	$s = j\omega$	$0 < \omega < \omega_0$	(11.5)
Path $\overline{bc}$:	$s = \lim_{\rho \to 0} \left(j\omega_0 + \rho e^{j\theta} \right)$	$-90° \le \theta \le 90°$	(11.6)
Path $\overline{cd}$:	$s = j\omega$	$\omega_0 \le \omega < \infty$	(11.7)
Path $\overline{def}$:	$s = \lim_{R \to \infty} Re^{j\theta}$	$+90° \le \theta \le -90°$	(11.8)
Path $\overline{fg}$:	$s = j\omega$	$-\infty < \omega < -\omega_0$	(11.9)
Path $\overline{gh}$:	$s = \lim_{\rho \to 0} \left(-j\omega_0 + \rho e^{j\theta} \right)$	$-90° \le \theta \le 90°$	(11.10)
Path $\overline{hi}$:	$s = j\omega$	$-\omega_0 < \omega < 0$	(11.11)
Path $\overline{ija}$:	$s = \lim_{\rho \to 0} \rho e^{j\theta}$	$-90° \le \theta \le 90°$	(11.12)

The **generalized Nyquist Path in the *z*-plane** is given in Fig. 11-13. Every pole and zero of $P(z)$ outside the unit circle is enclosed by the Nyquist Path when it is mapped into the $P(z)$-plane. In traversing the unit circle as a function of increasing angular frequency ω, any poles of $P(z)$ on the unit circle, which may include "integrators" at $z = 1$ (corresponding to $z \equiv e^{0 \cdot T} \equiv 1$ when $s = 0$), are excluded by infinitesimal circular arcs. For example, one pair of complex conjugate poles on the unit circle is shown in Fig. 11-13, circumvented by arcs of radius $\rho \to 0$. The remainder of the *z*-plane outside the unit circle is enclosed by the large circle of radius $R \to \infty$ shown in Fig. 11-13.

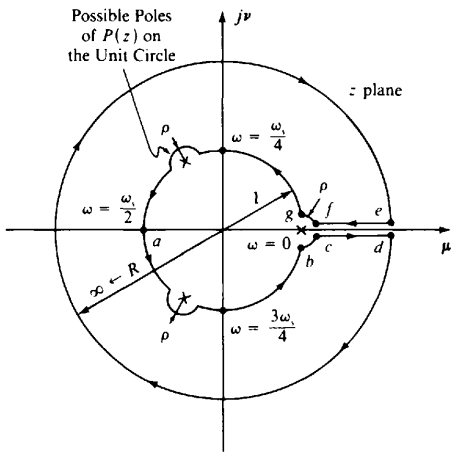


Fig. 11-13

The **unit circle in the *z*-plane** has a practical feature not shared by the Nyquist Path in the *s*-plane, one that facilitates drawing Polar Plots, as well as having other consequences in designing digital systems. First, we define the **angular sampling frequency** $\omega_s = 2\pi/T$ (radians per unit time). The advantage is that the unit circle repeats itself every angular sampling frequency ω_s as ω increases. This is shown in Fig. 11-14(*a*), which illustrates that the portion of the $j\omega$-axis in the *s*-plane between

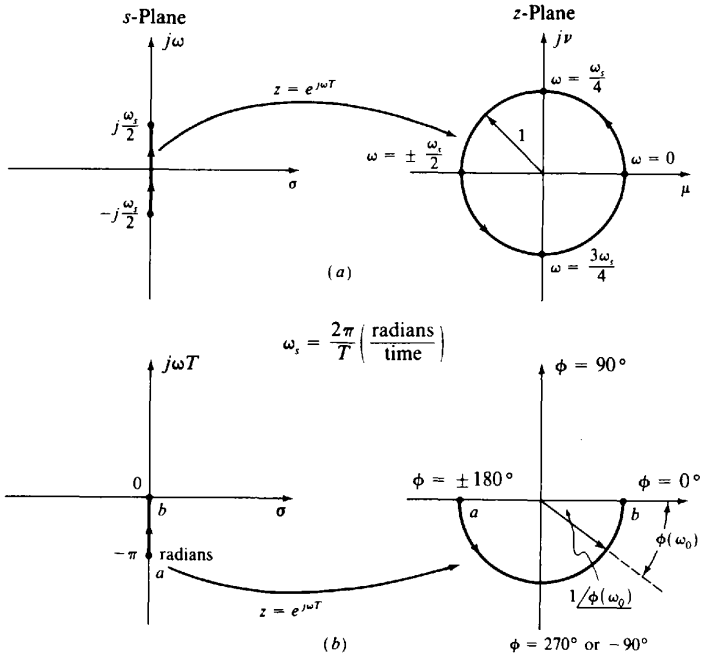


Fig. 11-14

$-j\omega_s/2$ and $+j\omega_s/2$ maps into the entire unit circle in the z-plane. This property is useful in drawing Polar Plots of functions $P(z) = P(e^{j\omega T})$, because the same Polar Plot is obtained for $n\omega_s \le \omega \le (n+1)\omega_s$, for any $n = \pm 1, \pm 2, \ldots$. Also, since the circular arc from $\omega = 0$ to $\omega_s/2$ is the mirror image of that from $\omega = -\omega_s/2$ to 0, the function $P(e^{j\omega T})$ need only be evaluated from $\omega = -\omega_s/2$ to 0 to obtain a complete Polar Plot, taking advantage of the symmetry of the mapping (Property 2, Section 11.6).

It is sometimes also convenient to treat the Polar Plot mapping as a function of ωT rather than ω. Then the strip $-(\omega_s/2)T \le \omega T \le 0$ is equivalent to $-\pi \le \omega T \le 0$ (in radians), because $\omega_s/2 = \pi/T$; this strip is mapped into the *lower half* of the unit circle in polar coordinates, from $-180°$ ($-\pi$ radians) to $0°$ or radians [Fig. 11-14(b)].

11.8 THE NYQUIST STABILITY PLOT

The Nyquist Stability Plot, an extension of the Polar Plot, is a mapping of the *entire* Nyquist Path into the P-plane. It is constructed using the mapping properties of Sections 11.4 and 11.6 and, for continuous systems, Equations (*11.5*) through (*11.8*) and Equation (*11.12*). A carefully drawn sketch is sufficient for most purposes.

A general construction procedure is outlined for continuous systems in the following steps.

Step 1: Check $P(s)$ for poles on the $j\omega$-axis and at the origin.

Step 2: Using Equation (*11.5*) through (*11.7*), sketch the image of path $\overline{ad}$ in the $P(s)$-plane. If there are no poles on the $j\omega$-axis, Equation (*11.6*) need not be employed. In this case, Step 2 should read: Sketch the Polar Plot of $P(j\omega)$.

Step 3: Draw the mirror image about the real axis Re P of the sketch resulting from Step 2. This is the mapping of path $\overline{fi}$.

Step 4: Use Equation (*11.8*) to plot the image of path $\overline{def}$. This path at infinity usually plots into a point in the $P(s)$-plane.

Step 5: Employ Equation (*11.12*) to plot the image of path $\overline{ija}$.

Step 6: Connect all curves drawn in the previous steps. Recall that the image of a closed contour is closed. The conformal mapping property helps by determining the image in the $P(s)$-plane of the corner angles of the semicircles in the Nyquist Path.

The procedure is similar for discrete-time systems, with the Nyquist Path given in Fig. 11-13 instead, as illustrated in Example 11.11 and Problems 11.65 through 11.72.

11.9 NYQUIST STABILITY PLOTS OF PRACTICAL FEEDBACK CONTROL SYSTEMS

For Nyquist stability analysis of linear feedback control systems, $P(\omega)$ is equal to the open-loop transfer function $GH(\omega)$. The most common control systems encountered in practice are those classified as type $0, 1, 2, \ldots, l$ systems (Chapter 9).

EXAMPLE 11.8. *Type 0 continuous system*

$$GH(s) = \frac{1}{s+1}$$

By definition, a *type 0 system* has no poles at the origin. This particular system has no poles on the $j\omega$-axis. The Nyquist Path is given in Fig. 11-15.

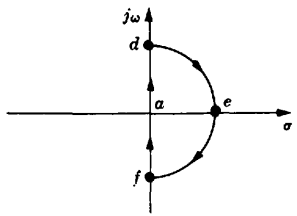


Fig. 11-15

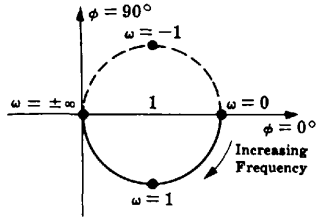


Fig. 11-16

The Polar Plot for this loop transfer function was constructed in Example 11.7, and it is shown in Fig. 11-16. This plot is the image of the $j\omega$-axis, or path $\overline{fad}$ of the Nyquist Path, in the $GH(s)$-plane. The semicircular path $\overline{def}$ at infinity is mapped into the $GH(s)$-plane in the following manner. Equation (*11.8*) implies substitution of $s = \lim_{R\to\infty} Re^{j\theta}$ into the expression for $GH(s)$, where $90° \le \theta \le -90°$. Hence

$$GH(s)|_{\text{path } \overline{def}} \equiv GH(\infty) = \frac{1}{\lim_{R\to\infty} Re^{j\theta} + 1}$$

By the elementary properties of limits,

$$GH(\infty) = \lim_{R\to\infty} \left[\frac{1}{Re^{j\theta} + 1} \right]$$

But since $|a+b| \ge |\,|a| - |b|\,|$, then

$$|GH(\infty)| = \lim_{R\to\infty} \left| \frac{1}{Re^{j\theta} + 1} \right| \le \lim_{R\to\infty} \left(\frac{1}{R-1} \right) = 0$$

and the infinite semicircle plots into a point at the origin. Of course, this computation was unnecessary for this simple example because the Polar Plot produces a completely closed contour in the $GH(s)$-plane. In fact, Polar Plots of all *type 0 systems* exhibit this property. The Nyquist Stability Plot is a replica of the Polar Plot with the axes relabeled, and is given in Fig. 11-17.

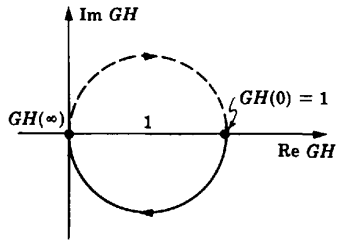


Fig. 11-17

EXAMPLE 11.9. *Type 1 continuous system*

$$GH(s) = \frac{1}{s(s+1)}$$

There is one pole at the origin. The Nyquist Path is given in Fig. 11-18.

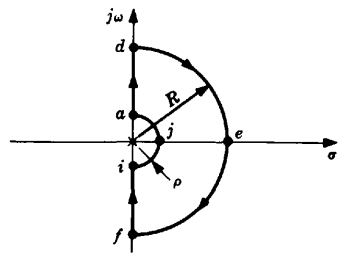


Fig. 11-18

Path $\overline{ad}$: $s = j\omega$ for $0 < \omega < \infty$, and

$$GH(j\omega) = \frac{1}{j\omega(j\omega+1)} = \frac{1}{\omega\sqrt{\omega^2+1}} \angle -90° - \tan^{-1}\omega$$

At extreme values of ω we have

$$\lim_{\omega\to 0} GH(j\omega) = \infty \angle -90° \qquad \lim_{\omega\to\infty} GH(j\omega) = 0 \angle -180°$$

As ω increases in the interval $0 < \omega < \infty$, the magnitude of GH decreases from ∞ to 0 and the phase angle decreases steadily from $-90°$ to $-180°$. Therefore the contour does not cross the negative real axis, but approaches it from below as shown in Fig. 11-19.

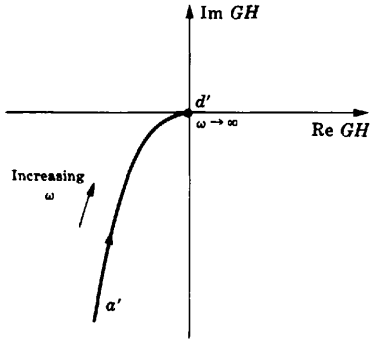


Fig. 11-19

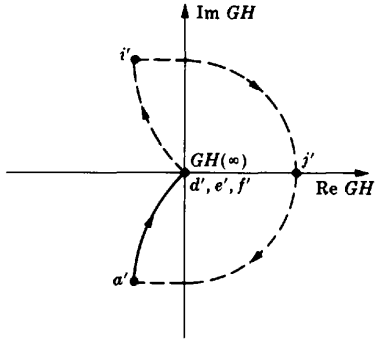


Fig. 11-20

Path $\overline{f'i'}$ is the mirror image about Re GH of path $\overline{a'd'}$. Since points d' and f' meet at the origin, the origin is clearly the image of path $\overline{def}$. Application of Equation (*11.8*) is therefore unnecessary.

Path $\overline{ija}$: $s = \lim_{\rho\to 0} \rho e^{j\theta}$ for $-90° \le \theta \le 90°$, and

$$\lim_{\rho\to 0} GH(\rho e^{j\theta}) = \lim_{\rho\to 0}\left[\frac{1}{\rho e^{j\theta}(\rho e^{j\theta}+1)}\right] = \lim_{\rho\to 0}\left[\frac{1}{\rho e^{j\theta}}\right] = \infty \cdot e^{-j\theta} = \infty \angle -\theta$$

where we have used the fact that $(\rho e^{j\theta} + 1) \to 1$ as $\rho \to 0$. Hence path $\overline{ija}$ maps into a semicircle of infinite radius. For point i, $GH = \infty \angle 90°$; for point j, $GH = \infty \angle 0°$; and for point a, $GH = \infty \angle -90°$. The resulting Nyquist Stability Plot is given in Fig. 11-20.

Path $\overline{i'j'a'}$ could also have been determined in the following manner. The Nyquist Path makes a 90° turn to the right at point i; hence by conformal mapping, a 90° right turn must be made at i' in the $GH(s)$-plane. The same goes for point a'. Since both i' and a' are points at infinity, and since the Nyquist Stability Plot must be a closed contour, a CW semicircle of infinite radius must join point i' to point a'.

Type *l* Continuous Systems

The Nyquist Stability Plot of a type l system includes l infinite semicircles in its path. There are $180l$ degrees in the connecting arc at infinity of the $GH(s)$-plane.

EXAMPLE 11.10. The type 3 system with

$$GH(s) = \frac{1}{s^3(s+1)}$$

has three infinite semicircles in its Nyquist Stability Plot (Fig. 11-21).

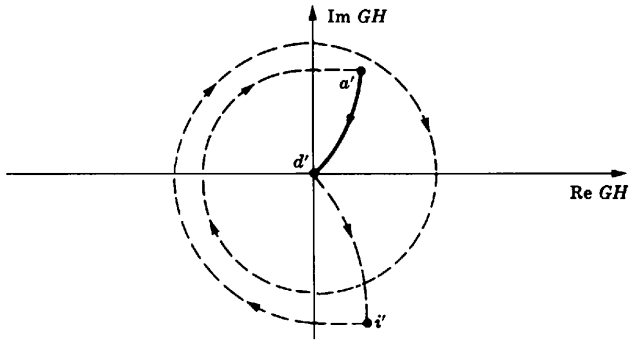


Fig. 11-21

Discrete-Time Systems

Nyquist Stability Plots of discrete-time systems are drawn in the same manner as above, the only difference being that the Nyquist Path is that given in Fig. 11-13, instead of Fig. 11-12.

EXAMPLE 11.11. Consider the type 1 digital control system with open-loop transfer function

$$GH(z) = \frac{K/4}{(z-1)(z-\frac{1}{2})}$$

The Polar Plot of GH is determined by first mapping the lower half of the unit circle in the z-plane into the GH-plane. This is readily accomplished with the aid of the mapping illustrated in Fig. 11-14(b), that is, we evaluate $GH(e^{j\omega T})$ for increasing values of ωT, from $-180°$ to 0° (or $-\pi$ to 0 radians). For given values of K and T, say $K = 1$ and $T = 1$,

$$GH(e^{j\omega T}) = \frac{K/4}{(e^{j\omega T}-1)(e^{j\omega T}-\frac{1}{2})} = \frac{1/4}{(e^{j\omega}-1)(e^{j\omega}-\frac{1}{2})}$$

For hand calculations, a combination of the Polar Form, Euler Form, and Complex Form are useful in evaluating $GH(e^{j\omega T})$ at different values of ω, because $e^{j\omega T} = 1 \angle \omega T(\text{rad}) = \cos(\omega T) + j\sin(\omega T) = \text{Re}(e^{j\omega T}) + j\,\text{Im}(e^{j\omega T})$. At $\omega = -\pi$ rad ($-180°$), we have

$$GH(e^{-j\pi}) = GH\left(1 \angle -180°\right) = \frac{0.25 \angle 0°}{\left(1 \angle -180° - 1 \angle 0°\right)\left(1 \angle -180° - \tfrac{1}{2} \angle 0°\right)}$$

$$= \frac{0.25}{(-1+j0-1)(-1+j0-\frac{1}{2})} = \frac{0.25}{(-2)(-\frac{3}{2})}$$

$$= 0.083 \angle 0°$$

Then, at $\omega = 270°$,

$$GH(e^{j3\pi/2}) = GH(e^{-j\pi/2}) = \frac{0.25\angle 0°}{(1\angle -90° - 1\angle 0°)(1\angle 90° - \frac{1}{2}\angle 0°)}$$

$$= \frac{0.25}{(-j-1)(-j-\frac{1}{2})} = \frac{0.25}{-\frac{1}{2}+j\frac{3}{2}}$$

$$= \frac{2(0.25)}{\sqrt{10}}\angle 180° - \tan^{-1}(3) \approx 0.158\angle -108.4°$$

Similarly, we find that $GH(e^{j2\pi})$ does not exist, but $\lim_{\phi \to 360°} GH(e^{j\phi}) = \lim_{\omega \to 0} GH(e^{j\omega}) = \infty\angle 90°$.

To complete the sketch of this half of the Polar Plot, we need to evaluate $GH(e^{j\omega})$ at a few more values of ω. We readily find $GH(e^{-j\pi/1000}) = 159\angle 90.5°$, $GH(e^{-j\omega/12}) = 1.8\angle 127°$, and $GH(e^{-j\pi/6}) = 0.779\angle 159°$. The result is shown as the dashed curve from a' to b' in Fig. 11-22, the mapping of a to b in Fig. 11-13. The remaining portion of the Polar Plot, for $\omega = 0$ to π, from g' to a' in Fig. 11-22, is the mirror image of a' to b' about the real axis, by Property 2 of Section 11.6. This portion, from g' to a', is drawn as a solid curve, keeping with the convention that the Polar Plot is highlighted for positive values of ω, $0 < \omega T < (2n-1)\pi$, $n = 1, 2, \ldots$.

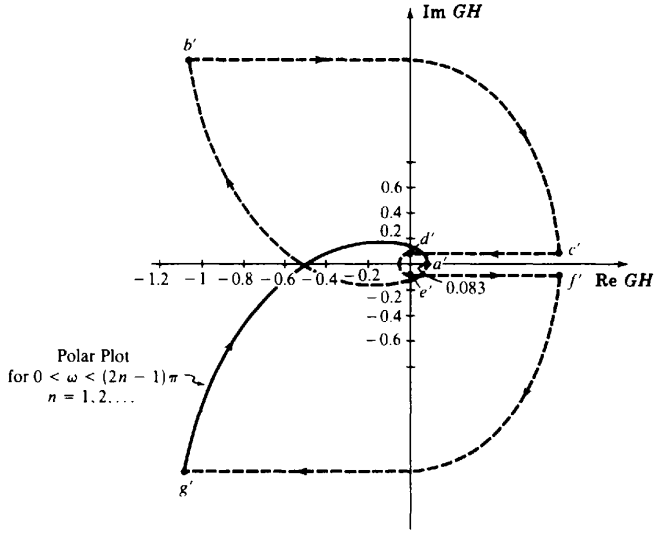


Fig. 11-22

The Nyquist Stability Plot is determined by completing the mapping of Fig. 11-13 segments b to c, c to d, d to e, and f to g, to the GH-plane. Using the mapping properties of Section 11.4 and limit calculations, $GH(e^{j\omega})$ makes a right turn at b', from $\infty\angle 90°$ to $\infty\angle 0°$ at c', then to $0\angle 0°$ at d' and at e', and $\infty\angle 0°$ at f' to $\infty\angle -90°$ at g', using limit operations for radii ρ and R in Fig. 11-13. For example, $\lim_{\rho \to 0} GH(z = 1 + \rho e^{j\theta})$ for $-90° < \theta < 0°$, provides the mapping of the arc from b to c in Fig. 11-13 into the arc from b' ($\infty\angle 90°$) to c' ($\infty\angle 0°$) in Fig. 11-22.

11.10 THE NYQUIST STABILITY CRITERION

A linear closed-loop *continuous control system* is absolutely stable if the roots of the characteristic equation have negative real parts (Section 5.2). Equivalently, the poles of the closed-loop transfer function, or the *zeros* of the denominator, $1 + GH(s)$, of the closed-loop transfer function, must lie in the left-half plane (LHP). For continuous systems, the Nyquist Stability Criterion establishes the number of *zeros* of $1 + GH(s)$ *in the RHP* directly from the Nyquist Stability Plot of $GH(s)$. For *discrete-time control systems*, the Nyquist Stability Criterion establishes the number of *zeros* of $1 + GH(z)$ *outside the unit circle* of the z-plane, the region of instability for discrete systems.

For either class of systems, continuous or discrete-time, the Nyquist Stability Criterion may be stated as follows.

Nyquist Stability Criterion

The closed-loop control system whose open-loop transfer function is GH is stable if and only if

$$N = -P_0 \le 0 \qquad (11.13)$$

where

$P_0 \equiv$ { *number of poles* (≥ 0) of GH in the RHP for continuous systems
number of poles (≥ 0) of GH outside the unit circle (of the z-plane) for discrete-time systems

$N \equiv$ total number of CW encirclements of the $(-1, 0)$ point (i.e., $GH = -1$) in the GH-plane (continuous or discrete)

If $N > 0$, *the number of zeros* Z_0 *of* $1 + GH$ *in the RHP* for continuous systems, or *outside the unit circle* for discrete systems, is determined by

$$Z_0 = N + P_0 \qquad (11.14)$$

If $N \le 0$, the $(-1, 0)$ point is not enclosed by the Nyquist Stability Plot. Therefore $N \le 0$ if the region to the right of the contour in the prescribed direction does not include the $(-1, 0)$ point. Shading this region helps significantly in determining whether $N \le 0$.

If $N \le 0$ *and* $P_0 = 0$, then the system is absolutely stable if and only if $N = 0$; that is, if and only if the $(-1, 0)$ point *does not* lie in the shaded region.

EXAMPLE 11.12. The Nyquist Stability Plot for $GH(s) = 1/s(s+1)$ was determined in Example 11.9 and is shown in Fig. 11-23. The region to the right of the contour has been shaded. Clearly, the $(-1, 0)$ point is not in the shaded region; therefore it is not enclosed by the contour and so $N \le 0$. The poles of $GH(s)$ are at $s = 0$ and $s = -1$, neither of which are in the RHP; hence $P_0 = 0$. Thus

$$N = -P_0 = 0$$

and the system is absolutely stable.

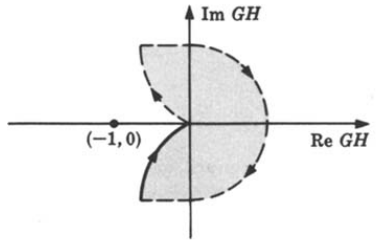


Fig. 11-23

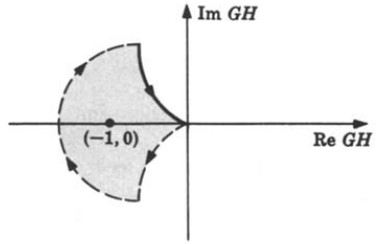


Fig. 11-24

EXAMPLE 11.13. The Nyquist Stability Plot for $GH(s) = 1/s(s-1)$ is given in Fig. 11-24. The region to the right of the contour has been shaded and the $(-1, 0)$ point is enclosed; then $N > 0$. (It is clear that $N = 1$.) The poles of GH are at $s = 0$ and $s = +1$, the latter pole being in the RHP. Hence $P_0 = 1$.

$N \neq -P_0$ indicates that the system is *unstable*. From Equation (*11.14*) we have

$$Z_0 = N + P_0 = 2$$

zeros of $1 + GH$ in the RHP.

EXAMPLE 11.14. The Nyquist Stability Plot for the discrete-time open-loop transfer function

$$GH(z) = \frac{K/4}{(z-1)(z-0.5)}$$

was determined in Example 11.11 and is repeated in Fig. 11-25 for $K = 1$. The region to the right of the contour has been shaded and the $(-1,0)$ point is not enclosed for $K = 1$. Thus $N \leq 0$ and from Equation (*11.13*) there are no poles outside the unit circle of the z-plane, that is, $P_0 = 0$. Hence $N = -P_0 = 0$ and the system is therefore stable.

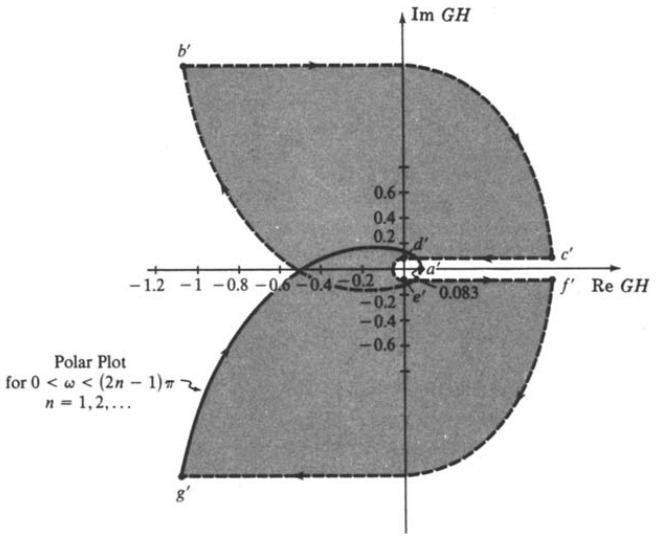


Fig. 11-25

11.11 RELATIVE STABILITY

The results in this section and the next are stated in terms of $GH(\omega)$, for either continuous $[GH(j\omega)]$ or discrete-time $[GH(e^{j\omega T})]$ systems.

The relative stability of a feedback control system is readily determined from the Polar or Nyquist Stability Plot.

The (angular) **phase crossover frequency** ω_π is that frequency at which the phase angle of $GH(\omega)$ is $-180°$, that is, the frequency at which the Polar Plot crosses the negative real axis. The **gain margin** is given by

$$\text{gain margin} = \frac{1}{|GH(\omega_\pi)|}$$

These quantities are illustrated in Fig. 11-26.

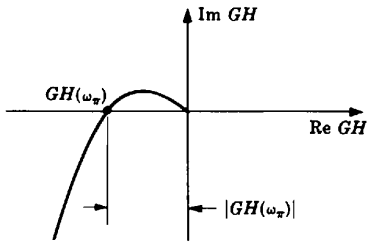


Fig. 11-26

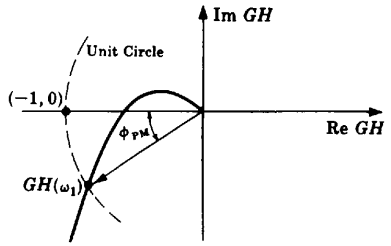


Fig. 11-27

The (angular) **gain crossover frequency** ω_1 is that frequency at which $|GH(\omega)| = 1$. The **phase margin** ϕ_{PM} is the angle by which the Polar Plot must be rotated to cause it to pass through the $(-1,0)$ point. It is given by

$$\phi_{PM} = [180 + \arg GH(\omega_1)] \text{ degrees}$$

These quantities are illustrated in Fig. 11-27.

11.12 *M*- AND *N*-CIRCLES*

The closed-loop frequency response of a unity feedback control system is given by

$$\frac{C}{R}(\omega) = \frac{G(\omega)}{1+G(\omega)} = \left|\frac{G(\omega)}{1+G(\omega)}\right| \angle \tan^{-1}\left[\frac{\text{Im}(C/R)(\omega)}{\text{Re}(C/R)(\omega)}\right] \qquad (11.15)$$

The magnitude and phase angle characteristics of the closed-loop frequency response of a unity feedback control system can be determined directly from the Polar Plot of $G(\omega)$. This is accomplished by first drawing lines of constant magnitude, called ***M*-circles**, and lines of constant phase angle, called ***N*-circles**, directly onto the $G(\omega)$-plane, where

$$M \equiv \left|\frac{G(\omega)}{1+G(\omega)}\right| \qquad (11.16)$$

$$N \equiv \frac{\text{Im}(C/R)(\omega)}{\text{Re}(C/R)(\omega)} \qquad (11.17)$$

The intersection of the Polar Plot with a particular M-circle yields the value of M at the frequency ω of $G(\omega)$ at the point of intersection. The intersection of the Polar Plot with a particular N-circle yields the value of N at the frequency ω of $G(\omega)$ at the intersection point. M versus ω and N versus ω are easily plotted from these points.

*The letter symbols M, N used in this section for M- and N-circles are not equal to and should not be confused with the manipulated variable $M = M(s)$ defined in Chapter 2 and with the number of encirclements N of the $(-1,0)$ point of Section 11.10. It is unfortunate that the same symbols have been used to signify more than one quantity. But in the interest of being consistent with most other control system texts, we have maintained the terminology of the classical literature and have now pointed this out to the reader.

Several M-circles are superimposed on a typical Polar Plot in the $G(\omega)$-plane in Fig. 11-28.

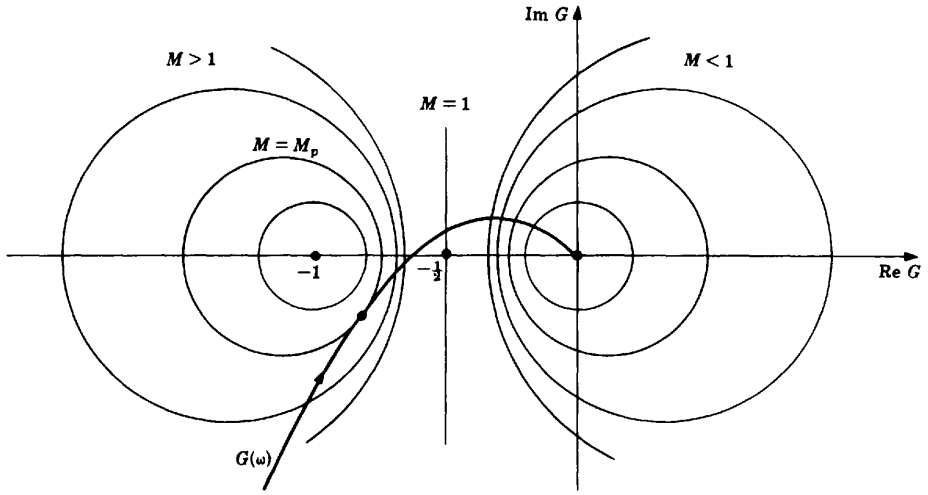


Fig. 11-28

The **radius of an M-circle** is given by

$$\text{radius of } M\text{-circle} = \left| \frac{M}{M^2 - 1} \right| \tag{11.18}$$

The **center of an M-circle** always lies on the Re $G(\omega)$-axis. The center point is given by

$$\text{center of } M\text{-circle} = \left(\frac{-M^2}{M^2 - 1}, 0 \right) \tag{11.19}$$

The **resonance peak** M_p is given by the largest value of M of the M-circle(s) tangent to the Polar Plot. (There may be more than one tangency.)

The **damping ratio** ζ for a second-order continuous system with $0 \le \zeta \le 0.707$ is related to M_p by

$$M_p = \frac{1}{2\zeta\sqrt{1 - \zeta^2}} \tag{11.20}$$

Several N-circles are superimposed on the Polar Plot shown in Fig. 11-29. The **radius of an N-circle** is given by

$$\text{radius of } N\text{-circle} = \sqrt{\frac{1}{4} + \left(\frac{1}{2N} \right)^2} \tag{11.21}$$

The **center of an N-circle** always falls on the line Re $G(\omega) = -\frac{1}{2}$. The center point is given by

$$\text{center of } N\text{-circle} = \left(-\frac{1}{2}, \frac{1}{2N} \right) \tag{11.22}$$

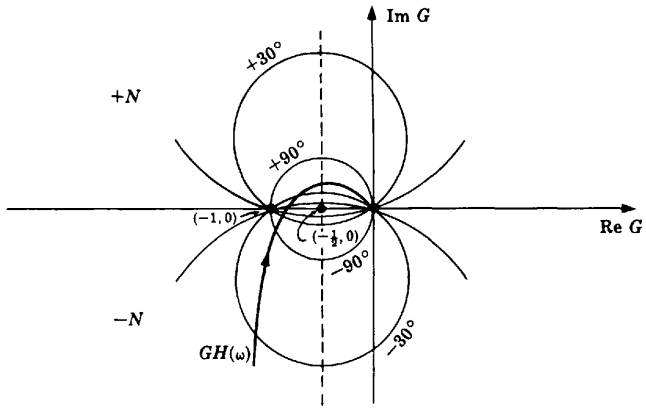


Fig. 11-29

Solved Problems

COMPLEX FUNCTIONS OF A COMPLEX VARIABLE

11.1. What are the values of $P(s) = 1/(s^2 + 1)$ for $s_1 = 2$, $s_2 = j4$, and $s_3 = 2 + j4$?

$$P(s_1) = P(2) = \frac{1}{(2)^2 + 1} = \frac{1}{5} + j0 \qquad P(s_2) = P(j4) = \frac{1}{(j4)^2 + 1} = -\frac{1}{15} + j0$$

$$P(s_3) = P(2 + j4) = \frac{1}{(2 + j4)^2 + 1} = \frac{1}{-11 + j16}$$

$$= \frac{1\angle 0°}{\sqrt{(11)^2 + (16)^2}\angle \tan^{-1}(16/-11)} = \frac{1}{19.4}\angle 0° - 124.6°$$

$$= 0.0514\angle -124.6° = -0.0514\angle 55.4° = -0.0292 - j0.0423$$

11.2. Map the imaginary axis in the s-plane onto the $P(s)$-plane, using the mapping function $P(s) = s^2$.

We have $s = j\omega$, $-\infty < \omega < \infty$. Therefore $P(j\omega) = (j\omega)^2 = -\omega^2$. Now when $\omega \to -\infty$, $P(j\omega) \to -\infty$ (or $-\infty^2$, if you prefer). When $\omega \to +\infty$, $P(j\omega) \to -\infty$; and when $\omega = 0$, $P(j0) = 0$. Thus as $j\omega$ increases along the negative *imaginary* axis from $-j\infty$ toward $j0$, $P(j\omega)$ increases along the negative *real* axis from $-\infty$ to 0. When $j\omega$ increases from $j0$ to $+j\infty$, $P(j\omega)$ decreases back to $-\infty$, again along the negative *real* axis. The mapping is plotted in the following manner (Fig. 11-30):

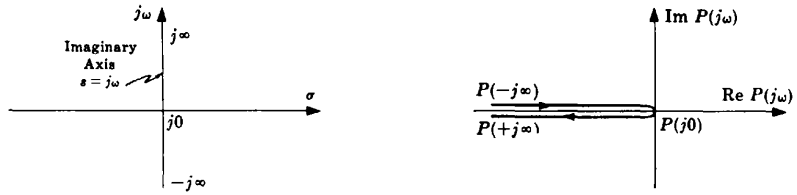


Fig. 11-30

The two lines in the $P(j\omega)$-plane are actually superimposed, but they are shown here separated for clarity.

11.3. Map the rectangular region in the s-plane bounded by the lines $\omega = 0$, $\sigma = 0$, $\omega = 1$, and $\sigma = 2$ onto the $P(s)$-plane using the transformation $P(s) = s + 1 - j2$.

We have

$$\omega = 0: \quad P(\sigma) = (\sigma + 1) - j2 \qquad \omega = 1: \quad P(\sigma + j1) = (\sigma + 1) - j1$$
$$\sigma = 0: \quad P(j\omega) = 1 + j(\omega - 2) \qquad \sigma = 2: \quad P(2 + j\omega) = 3 + j(\omega - 2)$$

Since σ varies over all real numbers $(-\infty < \sigma < \infty)$ on the line $\omega = 0$, so does $\sigma + 1$ on $P(\sigma) = (\sigma + 1) - j2$. Therefore $\omega = 0$ maps onto the line $-j2$ in the $P(s)$-plane. Similarly, $\sigma = 0$ maps onto the line $P(s) = 1$, $\omega = 1$ maps onto the line $P(s) = -j1$, and $\sigma = 2$ onto the line $P(s) = 3$. The resulting transformation is illustrated in Fig. 11-31.

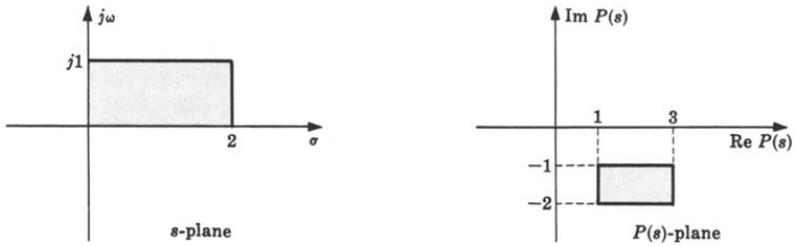


Fig. 11-31

This type of mapping is called a **translation** mapping. Note that the mapping would be exactly the same if $s = \sigma + j\omega$ were replaced by $z = \mu + j\nu$ in this example.

11.4. Find the derivative of $P(s) = s^2$ at the points $s = s_0$ and $s_0 = 1$.

$$\left.\frac{dP}{ds}\right|_{s=s_0} = \lim_{s \to s_0}\left[\frac{P(s) - P(s_0)}{s - s_0}\right] = \lim_{s \to s_0}\left[\frac{s^2 - s_0^2}{s - s_0}\right] = \lim_{s \to s_0}(s + s_0) = 2s_0$$

At $s_0 = 1$, we have $(dP/ds)|_{s=1} = 2$. Similarly, if $P(z) = z^2$, $(dP/dz)|_{z=1} = 2$.

ANALYTIC FUNCTIONS AND SINGULARITIES

11.5. Is $P(s) = s^2$ an analytic function in any region of the s-plane? If so, which region?

From the preceding problem $(dP/ds)|_{s=s_0} = 2s_0$. Hence s^2 is analytic wherever $2s_0$ is finite (Definition 11.1). Thus s^2 is analytic in the entire finite region of the s-plane. Such functions are often called **entire functions**. Similarly, z^2 is analytic in the entire finite region of the z-plane.

11.6. Is $P(s) = 1/s$ analytic in any region of the s-plane?

$$\left.\frac{dP}{ds}\right|_{s=s_0} = \lim_{s \to s_0}\left[\frac{1/s - 1/s_0}{s - s_0}\right] = \lim_{s \to s_0}\left[\frac{-(s - s_0)}{ss_0(s - s_0)}\right] = \frac{-1}{s_0^2}$$

This derivative is unique and finite for all $s_0 \neq 0$. Hence $1/s$ is analytic at all points in the s-plane except the origin, $s = s_0 = 0$. The point $s = 0$ is a *singularity* (pole) of $1/s$. Singularities other than poles exist, but not in the transfer functions of ordinary control system components.

11.7. Is $P(s) = |s|^2$ analytic in any region of the s-plane?

First put $s = \sigma + j\omega$, $s_0 = \sigma_0 + j\omega_0$. Then

$$\left.\frac{dP}{ds}\right|_{s=s_0} = \lim_{(s - s_0) \to 0}\left[\frac{|\sigma + j\omega|^2 - |\sigma_0 + j\omega_0|^2}{(\sigma + j\omega) - (\sigma_0 + j\omega_0)}\right]$$
$$= \lim_{[(\sigma - \sigma_0) + j(\omega - \omega_0)] \to 0}\left[\frac{(\sigma - \sigma_0)(\sigma + \sigma_0) + (\omega - \omega_0)(\omega + \omega_0)}{(\sigma - \sigma_0) + j(\omega - \omega_0)}\right]$$

If the limit exists it must be unique and should not depend on how s approaches s_0, or equivalently how $[(\sigma - \sigma_0) + j(\omega - \omega_0)]$ approaches zero. So first let $s \to s_0$ along the $j\omega$-axis and obtain

$$\left.\frac{dP}{ds}\right|_{s=s_0} = \lim_{\substack{\omega \to \omega_0 \\ \sigma = \sigma_0}}\left[\frac{(\omega - \omega_0)(\omega + \omega_0)}{j(\omega - \omega_0)}\right] = -j2\omega_0$$

Now let $s \to s_0$ along the σ-axis; that is,

$$\left.\frac{dP}{ds}\right|_{s=s_0} = \lim_{\substack{\sigma \to \sigma_0 \\ \omega = \omega_0}}\left[\frac{(\sigma - \sigma_0)(\sigma + \sigma_0)}{\sigma - \sigma_0}\right] = 2\sigma_0$$

Hence the limit *does not* exist for arbitrary nonzero values of σ_0 and ω_0, and therefore $|s|^2$ is not analytic anywhere in the s-plane except possibly at the origin. When $s_0 = 0$,

$$\left.\frac{dP}{ds}\right|_{s=0} = \lim_{s \to 0}\left[\frac{|s|^2 - 0}{s}\right] = \lim_{s \to 0}\left[\frac{(\sigma + j\omega)(\sigma - j\omega)}{\sigma + j\omega}\right] = 0$$

Therefore $P(s) = |s|^2$ is analytic only at the origin, $s = 0$.

11.8. If $P(s)$ is analytic at s_0, prove that it must be continuous at s_0. That is, show that $\lim_{s \to s_0} P(s) = P(s_0)$.

Since

$$P(s) - P(s_0) = \frac{P(s) - P(s_0)}{(s - s_0)} \cdot (s - s_0)$$

for $s \neq s_0$, then

$$\lim_{s \to s_0}[P(s) - P(s_0)] = \lim_{s \to s_0}\left[\frac{P(s) - P(s_0)}{(s - s_0)}\right] \cdot \lim_{s \to s_0}(s - s_0) = \left[\left.\frac{dP}{ds}\right|_{s=s_0}\right] \cdot 0 = 0$$

because $(dP/ds)|_{s=s_0}$ exists by hypothesis [i.e., $P(s)$ is analytic]. Therefore

$$\lim_{s \to s_0}[P(s) - P(s_0)] = 0 \qquad \text{or} \qquad \lim_{s \to s_0} P(s) = P(s_0)$$

11.9. **Polynomial functions** are defined by $Q(s) \equiv a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$, where $a_n \neq 0$, n is a positive integer called the **degree of the polynomial**, and $a_0, a_1, \ldots, a_n$ are constants. Prove that $Q(s)$ is analytic in every bounded (finite) region of the s-plane.

First consider s^n:

$$\left.\frac{d}{ds}[s^n]\right|_{s=s_0} = \lim_{s \to s_0}\left[\frac{s^n - s_0^n}{s - s_0}\right] = \lim_{s \to s_0}\left(s^{n-1} + s^{n-2}s_0 + \cdots + ss_0^{n-2} + s_0^{n-1}\right) = ns_0^{n-1}$$

Thus s^n is analytic in every finite region of the s-plane. Then, by mathematical induction, s^{n-1}, $s^{n-2}, \ldots, s$ are also analytic. Hence, by the elementary theorems on limits of sums and products, we see that $Q(s)$ is analytic in every finite region of the s-plane.

11.10. **Rational algebraic functions** are defined by $P(s) \equiv N(s)/D(s)$, where $N(s)$ and $D(s)$ are polynomials. Show that $P(s)$ is analytic at every point s where $D(s) \neq 0$; that is, prove that the transfer functions of control system elements that take the form of rational algebraic functions are analytic except at their poles.

The overwhelming majority of linear control system elements are in this category. The fundamental theorem of algebra, "a polynomial of degree n has n zeros and can be expressed as a product of n linear factors," helps to put $P(s)$ in a form more recognizable as a control system transfer function; that is, $P(s)$ can be written in the familiar form

$$P(s) \equiv \frac{N(s)}{D(s)} = \frac{b_m s^m + b_{m-1}s^{m-1} + \cdots + b_0}{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_0} = \frac{b_m(s+z_1)(s+z_2)\cdots(s+z_m)}{a_n(s+p_1)(s+p_2)\cdots(s+p_n)}$$

where $-z_1, -z_2, \ldots, -z_n$ are zeros, $-p_1, -p_2, \ldots, -p_n$ are poles, and $m \leq n$.

From the identity given by

$$\frac{N(s)}{D(s)} - \frac{N(s_0)}{D(s_0)} \equiv \frac{1}{D(s)D(s_0)}\left[D(s_0)(N(s) - N(s_0)) - N(s_0)(D(s) - D(s_0))\right]$$

where $D(s) \neq 0$, we get

$$\begin{aligned}
\left.\frac{dP}{ds}\right|_{s=s_0} &= \lim_{s\to s_0}\left[\frac{\dfrac{N(s)}{D(s)} - \dfrac{N(s_0)}{D(s_0)}}{s - s_0}\right] \\
&= \lim_{s\to s_0}\left[\frac{1}{D(s)D(s_0)}\left(D(s_0)\left[\frac{N(s)-N(s_0)}{s-s_0}\right] - N(s_0)\left[\frac{D(s)-D(s_0)}{s-s_0}\right]\right)\right] \\
&= \lim_{s\to s_0}\left[\frac{1}{D(s)}\left(\frac{N(s)-N(s_0)}{s-s_0}\right)\right] - \lim_{s\to s_0}\left[\frac{N(s_0)}{D(s)D(s_0)}\left(\frac{D(s)-D(s_0)}{s-s_0}\right)\right] \\
&= \lim_{s\to s_0}\left[\frac{1}{D(s)}\right]\cdot\lim_{s\to s_0}\left[\frac{N(s)-N(s_0)}{s-s_0}\right] - \lim_{s\to s_0}\left[\frac{N(s_0)}{D(s)D(s_0)}\right]\cdot\lim_{s\to s_0}\left[\frac{D(s)-D(s_0)}{s-s_0}\right] \\
&= \frac{1}{D(s_0)}\cdot\left.\frac{dN}{ds}\right|_{s=s_0} - \frac{N(s_0)}{D(s_0)^2}\cdot\left.\frac{dD}{ds}\right|_{s=s_0}
\end{aligned}$$

where we have used the results of Problems 11.8, 11.9, and Definition 11.1. Therefore the derivative of $P(s)$ exists ($P(s)$ is analytic) for all points s where $D(s) \neq 0$.

Note that we have determined a formula for the derivative of a rational algebraic function (the last part of the above equation) in terms of the derivatives of its numerator and denominator, in addition to solving the required problem.

11.11. Prove that e^{-sT} is analytic in every bounded region of the s-plane.

In complex variable theory e^{-sT} is defined by the power series

$$e^{-sT} = \sum_{k=0}^{\infty}\frac{(-sT)^k}{k!}$$

By the ratio test, as $k \to \infty$ we have

$$\left|\frac{(-sT)^k/k!}{(-sT)^{k+1}/(k+1)!}\right| = \left|\frac{k+1}{-sT}\right| \to \infty$$

Hence the radius of convergence of this power series is infinite. The sum of a power series is analytic within its radius of convergence. Thus e^{-sT} is analytic in every bounded region of the s-plane.

11.12. Prove that $e^{-sT}P(s)$ is analytic wherever $P(s)$ is analytic. Hence systems containing a combination of rational algebraic transfer functions and time-delay operators (i.e., e^{-sT}) are analytic except at the poles of the system.

By Problem 11.11, e^{-sT} is analytic in every bounded region of the s-plane; and by Problem 11.10, $P(s)$ is analytic except at its poles. Now

$$\begin{aligned}
\left.\frac{d}{ds}\left[e^{-sT}P(s)\right]\right|_{s=s_0} &= \lim_{s\to s_0}\left[\frac{e^{-sT}P(s) - e^{-s_0T}P(s_0)}{s - s_0}\right] \\
&= \lim_{s\to s_0}\left[e^{-sT}\left(\frac{P(s)-P(s_0)}{s-s_0}\right) + P(s_0)\left(\frac{e^{-sT}-e^{-s_0T}}{s-s_0}\right)\right] \\
&= e^{-s_0T}\left.\frac{dP}{ds}\right|_{s=s_0} + P(s_0)\left.\frac{d}{ds}(e^{-sT})\right|_{s=s_0}
\end{aligned}$$

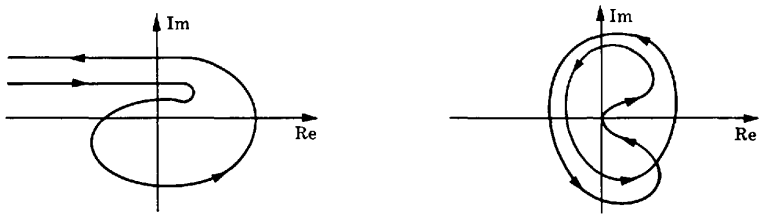
Therefore $e^{-sT}P(s)$ is analytic wherever $P(s)$ is analytic.

11.13. Consider the function given by $P(s) = e^{-sT}(s^2 + 2s + 3)/(s^2 - 2s + 2)$. Where are the singularities of this function? Where is $P(s)$ analytic?

The singular points are at the poles of $P(s)$. Since $s^2 - 2s + 2 = (s - 1 + j1)(s - 1 - j1)$, the two poles are given by $-p_1 = 1 - j1$ and $-p_2 = 1 + j1$. $P(s)$ is analytic in every bounded region of the s-plane except at the points $s = -p_1$ and $s = -p_2$.

CONTOURS AND ENCIRCLEMENTS

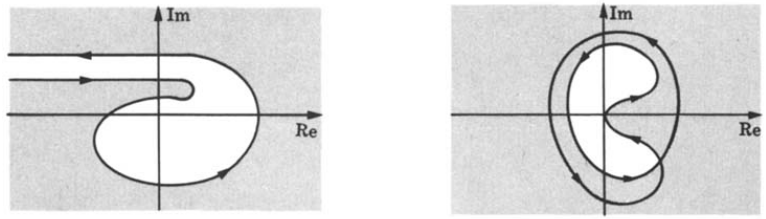
11.14. What points are *enclosed* by the following contours (Fig. 11-32)?



(a) (b)

Fig. 11-32

By shading the region to the right of each contour as it is traversed in the prescribed direction, we get Fig. 11-33. All points in the shaded regions are enclosed.



(a) (b)

Fig. 11-33

11.15. What contours of Problem 11.14 are *closed*?

Clearly, the contour of part (b) is closed. The contour of part (a) may or may not close upon itself at infinity in the complex plane. This cannot be determined from the given graph.

11.16. What is the *direction* (positive or negative) of each contour in Problem 11.14(a) and (b)?

Using the origin as a base, each contour is directed in the counterclockwise, *negative* direction about the origin.

11.17. Determine the number of encirclements N_0 of the origin for the contour in Fig. 11-34.

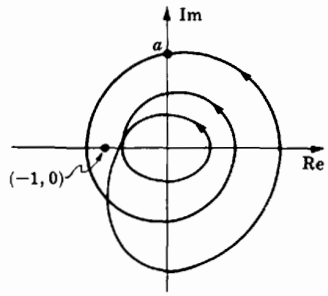


Fig. 11-34

Beginning at the point a, we rotate a radial line from the origin to the contour in the direction of the arrows. Three counterclockwise rotations of 360° result in the radial line returning to the point a. Hence $N_0 = -3$.

11.18. Determine the number of encirclements N_0 of the origin for the contour in Fig. 11-35.

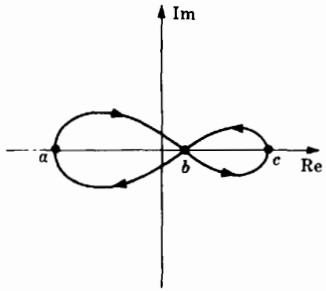


Fig. 11-35

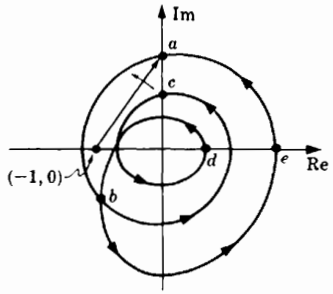


Fig. 11-36

Beginning at point a, $+180°$ is swept out by the contour when b is reached for the first time. In going from b to c and back to b, the net angular gain is zero. Returning to a from b yields $+180°$. Thus $N_0 = +1$.

11.19. Determine the number of encirclements N of the $(-1,0)$ point (i.e., the -1 point on the real axis) for the contour of Problem 11.17.

Again beginning at point a, we rotate a radial line *from the* $(-1,0)$ *point* to the contour in the direction of the arrows as shown in Fig. 11-36. In going from a to b to c, the radial line sweeps out somewhat less than $-360°$. But from c to d and back to b, the angle increases again toward the value reached in going only from a to b. Then from b to e to a the resultant angle is $-360°$. Thus $N = -1$.

PROPERTIES OF THE MAPPING P

11.20. Are the following functions single-valued: (a) $P(s) = s^2$, (b) $P(s) = s^{1/2}$?

(a) Substitution of any complex number s into $P(s) = s^2$ yields a unique value for $P(s)$. Hence $P(s) = s^2$ is a single-valued function.

(b) In polar form we have $s = |s|e^{j\theta}$, where $\theta = \arg(s)$. Therefore $s^{1/2} = |s|^{1/2}e^{j\theta/2}$. Now if we increase θ by 2π we return to the same point s. But

$$P(s) = |s|^{1/2}e^{j(\theta+2\pi)/2} = |s|^{1/2}e^{j\theta/2}e^{j\pi} = P(s)e^{j\pi}$$

which is *another* point in the $P(s)$-plane. Hence $P(s) = s^{1/2}$ has two points in the $P(s)$-plane for every point in the s-plane. It is not a single-valued function; it is a **multiple-valued function** (with two values).

11.21. Prove that every closed contour containing no singular points of $P(s)$ in the s-plane maps into a closed contour in the $P(s)$-plane.

Suppose not. Then at some point s_0 where the s-plane contour closes upon itself the $P(s)$-plane contour is not closed. This means that one (nonsingular) point s_0 in the s-plane is mapped into more than one point in the $P(s)$-plane (the images of the point s_0). This contradicts the fact that $P(s)$ is a single-valued function (Property 1, Section 11.4).

11.22. Prove that P is a conformal mapping wherever P is analytic and $dP/ds \neq 0$.

Consider two curves: C in the s-plane and C', the image of C, in the $P(s)$-plane. Let the curve in the s-plane be described by a parameter t; that is, each t corresponds to a point $s = s(t)$ along the curve C. Hence C' is described by $P[s(t)]$ in the $P(s)$-plane. The derivatives ds/dt and dP/dt represent tangent vectors to corresponding points on C and C'. Now

$$\left.\frac{dP[s(t)]}{dt}\right|_{P(s)=P(s_0)} = \left.\frac{ds}{dt}\cdot\frac{dP(s)}{ds}\right|_{s=s_0}$$

where we have used the fact that P is analytic at some point $s_0 \equiv s(t_0)$. Put $dP/dt \equiv r_1 e^{j\phi}$, $dP/ds \equiv r_2 e^{j\alpha}$, and $ds/dt \equiv r_3 e^{j\theta}$. Then

$$r_1(s_0)e^{j\phi(s_0)} = r_2(s_0)\cdot r_3(s_0)e^{j[\theta(s_0)+\alpha(s_0)]}$$

Equating angles, we have $\phi(s_0) = \theta(s_0) + \alpha(s_0) = \theta(s_0) + \arg(dP/ds)|_{s=s_0}$, and we see that the tangent to C at s_0 is rotated through an angle $\arg(dP/ds)|_{s=s_0}$ at $P(s_0)$ on C' in the $P(s)$-plane.

Now consider two curves C_1 and C_2 intersecting at s_0, with images C_1' and C_2' in the $P(s)$-plane (Fig. 11-37).

Let θ_1 be the angle of inclination of the tangent to C_1, and θ_2 for C_2. Then the angles of inclination for C_1' and C_2' are $\theta_1 + \arg(dP/ds)|_{s=s_0}$, and $\theta_2 + \arg(dP/ds)|_{s=s_0}$. Therefore the angle $(\theta_1 - \theta_2)$ between C_1 and C_2 is equal in magnitude and sense to the angle between C_1' and C_2',

$$\theta_1 + \arg\left.\frac{dP}{ds}\right|_{s=s_0} - \theta_2 - \arg\left.\frac{dP}{ds}\right|_{s=s_0} = \theta_1 - \theta_2$$

Note that $\arg(dP/ds)|_{s=s_0}$ is indeterminate if $(dP/ds)|_{s=s_0} = 0$.

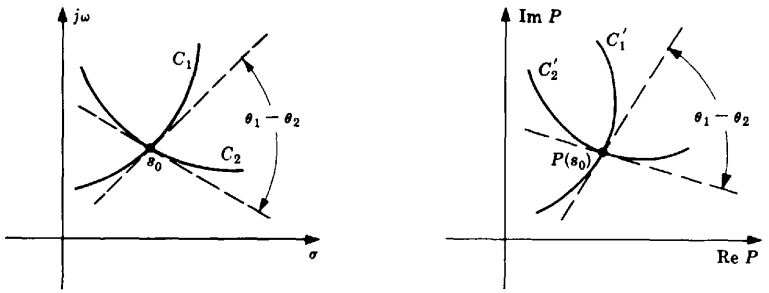


Fig. 11-37

11.23. Show that $P(s) = e^{-sT}$ is conformal in every bounded region of the s-plane.

e^{-sT} is analytic (Problem 11.11). Moreover, $(d/ds)(e^{-sT}) = -Te^{-sT} \neq 0$ in any bounded (finite) region of the s-plane. Then by Problem 11.22, $P(s) = e^{-sT}$ is conformal.

11.24. Show that $P(s)e^{-sT}$ is conformal for rational $P(s)$ and $dP/ds \neq 0$.

By Problem 11.12, Pe^{-sT} is analytic except at the poles of P. By Problem 11.12,

$$\frac{d}{ds}[Pe^{-sT}] = e^{-sT}\frac{dP}{ds} - PTe^{-sT} = e^{-sT}\left(\frac{dP}{ds} - TP\right)$$

Suppose $(d/ds)[Pe^{-sT}] = 0$. Then since $e^{-sT} \neq 0$ for any finite s, we have $dP/ds - TP = 0$ whose general solution is $P(s) = ke^{sT}$, k constant. But P is rational and e^{sT} is not. Hence $(d/ds)[Pe^{-sT}] \neq 0$.

11.25. Two s-plane contours C_1 and C_2 intersect in a 90° angle in Fig. 11-38. The analytic function $P(s)$ maps these contours into the $P(s)$-plane and $dP/ds \neq 0$ at s_0. Sketch the image of contour C_2 in a neighborhood of $P(s_0)$. The image of C_1 is also given.

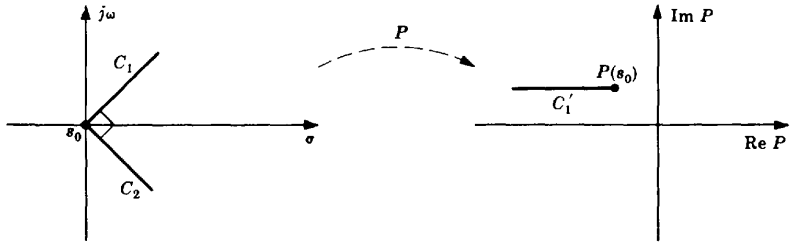


Fig. 11-38

By Problem 11.22, P is conformal; hence the angle between C_1' and C_2' is 90°. Since C_1 makes a left turn onto C_2 at s_0, then C_1' must also turn left at $P(s_0)$ (Fig. 11-39).

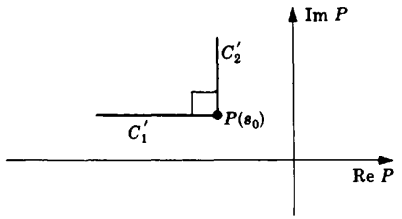


Fig. 11-39

11.26. Prove Equation (*11.1*): $N_0 = Z_0 - P_0$.

The bulk of the proof is somewhat more involved than can be handled with the complex-variable theory presented in this book. So we assume knowledge of a well-known theorem of functions of a complex variable and continue from there. The theorem states that if C is a closed contour in the s-plane, $P(s)$ an analytic function on C and within C except for possible poles, and $P(s) \neq 0$ on C, then

$$\frac{1}{2\pi j}\int_C \frac{P'(s)}{P(s)}\,ds = Z_0 - P_0$$

where Z_0 is the total number of zeros inside C, P_0 the total number of poles inside C, and $P' \equiv dP/ds$. Multiple poles and zeros are counted one for one; that is, a double pole at a point is two poles of the total, a triple zero is three zeros of the total.

Now since $d[\ln P(s)] = [P'(s)/P(s)]\,ds$ and $\ln P(s) \equiv \ln|P(s)| + j \arg P(s)$, we have

$$\frac{1}{2\pi j}\int_C\left[\frac{P'(s)}{P(s)}\right]ds = \frac{1}{2\pi j}\int_C d[\ln P(s)] = \frac{1}{2\pi j}[\ln P(s)]\Big|_C = \frac{1}{2\pi j}[\ln|P(s)| + j\arg P(s)]\Big|_C$$

$$= \frac{1}{2\pi j}[\ln|P(s)|]\Big|_C + \frac{1}{2\pi j}[j\arg P(s)]\Big|_C$$

Now since $\ln|P(s)|$ returns to its original value when we go once around C, the first term in the last equation is zero. Hence

$$Z_0 - P_0 = \frac{1}{2\pi}[\arg P(s)]\Big|_C$$

Since C is closed, the image of C in the $P(s)$-plane is closed, and the net change in the angle $\arg P(s)$ around the $P(s)$ contour is 2π times the number of encirclements N_0 of the origin in the $P(s)$-plane. Then $Z_0 - P_0 = 2N_0\pi/2\pi = N_0$. This result is often called *the principle of the argument*. Note that this result would be the same if we replaced s by z in all of the above. Therefore Equation (*11.1*) is valid for discrete-time systems as well.

11.27. Determine the number N_0 of P-plane contour encirclements for the complex-plane contour mapped into the P-plane shown in Fig. 11-40.

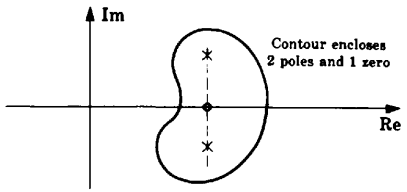


Fig. 11-40

$P_0 = 2$, $Z_0 = 1$. Therefore $N_0 = 1 - 2 = -1$.

11.28. Determine the number of zeros Z_0 enclosed by the complex-plane contour in Fig. 11-41, where $P_0 = 5$.

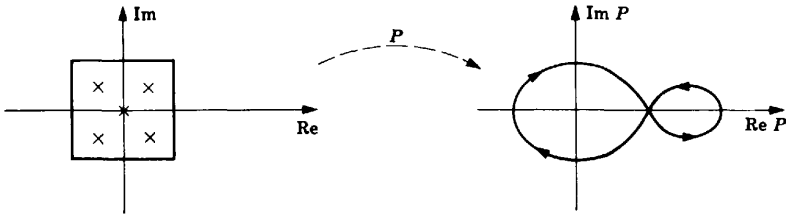


Fig. 11-41

$N_0 = 1$ was computed in Problem 11.18 for the given P-plane contour. Since $P_0 = 5$, then $Z_0 = N_0 + P_0 = 1 + 5 = 6$.

11.29. Determine the number of poles P_0 enclosed by the complex-plane contour in Fig. 11-42, where $Z_0 = 0$.

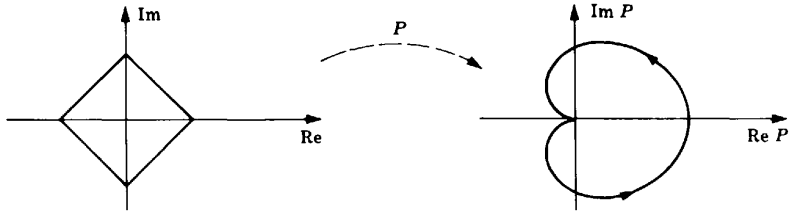


Fig. 11-42

Clearly, $N_0 = -1$. Hence $P_0 = Z_0 - N_0 = 0 + 1 = 1$.

11.30. Determine N_0 [Equation (*11.1*)] for the transfer function (transformation) and s-plane contour of Fig. 11-43.

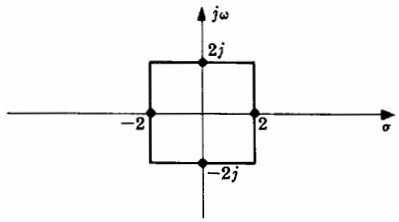


Fig. 11-43

The pole-zero map of $P(s)$ is given in Fig. 11-44. Hence three poles (two at $s = 0$ and one at $s = -1$) and no zeros are enclosed by the contour. Thus $P_0 = 3$, $Z_0 = 0$, and $N_0 = -3$.

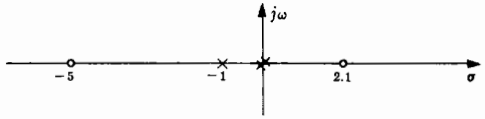


Fig. 11-44

11.31. Is the origin *enclosed* by the contour in Fig. 11-45?

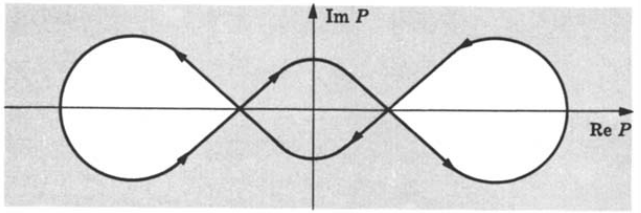


Fig. 11-45

The region to the right of the contour has been shaded. The origin falls in a shaded region and is therefore *enclosed* by the contour.

11.32. What is the sign of N_0 in Problem 11.31?

Since the origin is enclosed by the contour in a clockwise direction, $N_0 > 0$.

POLAR PLOTS

11.33. Prove Property 1 of Section 11.6.

Let $P(\omega) \equiv P_1(\omega) + jP_2(\omega)$ and $a \equiv a_1 + ja_2$, where $P_1(\omega)$, $P_2(\omega)$, a_1, and a_2 are real. Then

$$P(\omega) + a = (P_1(\omega) + a_1) + j(P_2(\omega) + a_2)$$

and the image of any point $(P_1(\omega), P_2(\omega))$ in the $P(\omega)$-plane is $(P_1(\omega) + a_1, P_2(\omega) + a_2)$ in the

$(P(\omega)+a)$-plane. Hence the image of a $P(\omega)$ contour is simply a *translation* (see Problem 11.3). Clearly, translation of the contour by a units is equivalent to translation of the axes (origin) by $-a$ units.

11.34. Prove Property 2 of Section 11.6.

The transfer function $P(s)$ of a constant-coefficient linear system is, in general, a ratio of polynomials with constant coefficients. The complex roots of such polynomials occur in conjugate pairs; that is, if $a+jb$ is a root, then $a-jb$ is also a root. If we let an asterisk (*) represent complex conjugation, then $a+jb=(a-jb)^*$, and if $a=0$, then $jb=(-jb)^*$. Therefore $P(j\omega)=P(-j\omega)^*$ or $P(-j\omega)=P(j\omega)^*$. Graphically this means that the plot for $P(-j\omega)$ is the mirror image about the real axis of the plot for $P(j\omega)$ since only the imaginary part of $P(j\omega)$ changes sign.

11.35. Sketch the Polar Plot of each of the following complex functions:

(a) $P(j\omega)=\omega^2\angle 45°$, (b) $P(j\omega)=\omega^2(\cos 45°+j\sin 45°)$, (c) $P(j\omega)=0.707\omega^2+0.707j\omega^2$.

(a) $\omega^2\angle 45°$ is in the form of Equation (*11.2*). Hence polar coordinates are used in Fig. 11-46.

(b) $P(j\omega)=\omega^2(\cos 45°+j\sin 45°)=\omega^2(0.707+0.707j)$

That is, $P(j\omega)$ is in the form of Equation (*11.3*) or (*11.4*). Hence rectangular coordinates is the natural choice as shown in Fig. 11.47.

Note that this graph is identical with that of part (a) except for the coordinates. In fact, $\omega^2(0.707+0.707j)=\omega^2\angle 45°$.

(c) Clearly, (c) is identical with (b), and therefore with (a). Among other things, this problem has illustrated how a complex function of frequency ω can be written in three different but mathematically and graphically identical forms: the polar form, Equation (*11.2*); the trigonometric or *Euler form*, Equation (*11.3*); and the equivalent rectangular (complex) form, Equation (*11.4*).

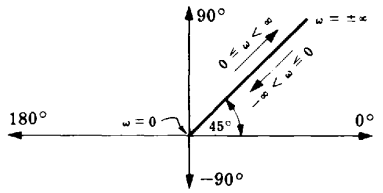


Fig. 11-46

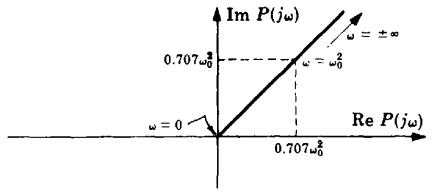


Fig. 11-47

11.36. Sketch the Polar Plot of

$$P(j\omega)=0.707\omega^2(1+j)+1$$

The Polar Plot of $0.707\omega^2(1+j)$ was drawn in Problem 11.35(b). By Property 1 of Section 11.6, the required Polar Plot is given by that of Problem 11.35(b) with its origin shifted to $-a=-1$ as shown in Fig. 11-48.

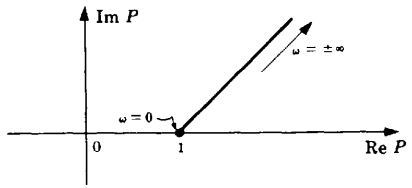


Fig. 11-48

11.37. Construct a Polar Plot from the set of graphs of the magnitude and phase angle of $P(j\omega)$ in Fig. 11-49, representing the frequency response of a linear constant-coefficient system.

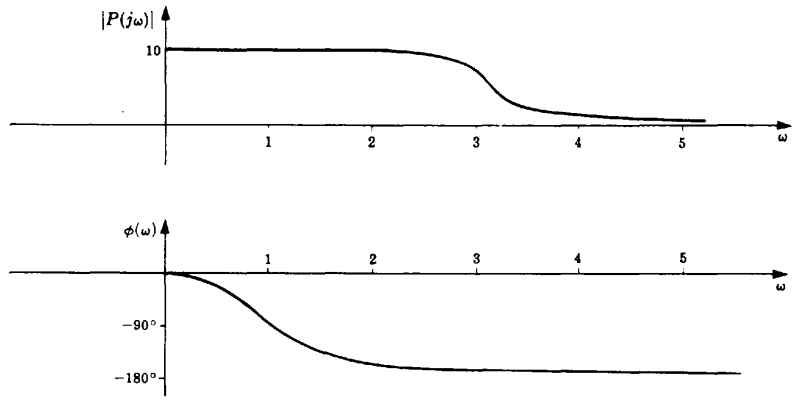


Fig. 11-49

The graphs shown above differ little from *Bode representations*, discussed in detail in Chapter 15. The Polar Plot is constructed by mapping this set of graphs into the $P(j\omega)$-plane. It is only necessary to choose values of ω and corresponding values of $|P(j\omega)|$ and $\phi(\omega)$ from the graphs and plot these points in the $P(j\omega)$-plane. For example at $\omega=0$, $|P(j\omega)|=10$ and $\phi(\omega)=0$. The resulting Polar Plot is given in Fig. 11-50.

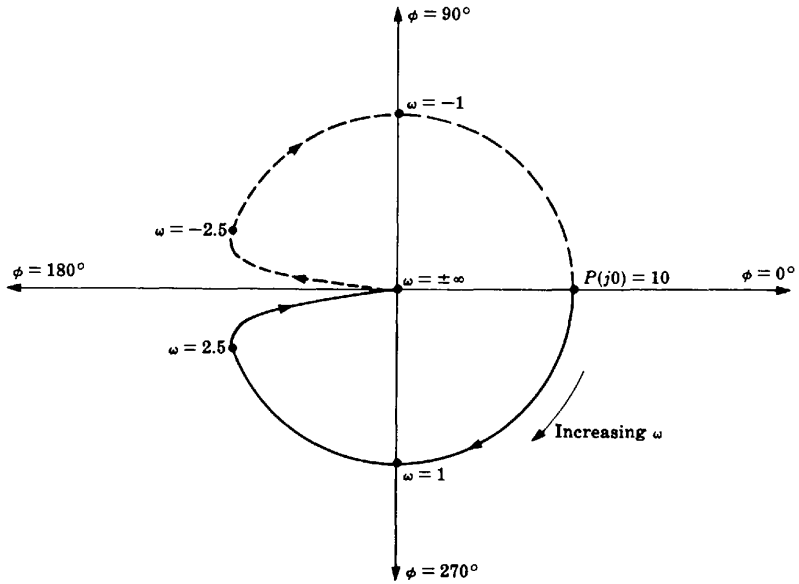


Fig. 11-50

The portion of the graph for $-\infty < \omega < 0$ has been drawn using the property of conjugate symmetry (Section 11.6).

11.38. Sketch the Polar Plot for

$$GH(s) = \frac{1}{s^4(s+p)} \qquad p > 0$$

Substituting $j\omega$ for s, and applying Equation (*11.2*), we obtain

$$GH(j\omega) = \frac{1}{j^4\omega^4(j\omega + p)}$$

$$= \frac{1}{\omega^4\sqrt{\omega^2 + p^2}} \angle -\tan^{-1}(\omega/p)$$

For $\omega = 0$ and $\omega \to \infty$, we have

$$GH(j0) = \infty \angle 0^\circ \qquad \lim_{\omega \to \infty} GH(j\omega) = 0 \angle -90^\circ$$

Clearly, as ω increases from zero to infinity, the phase angle remains negative and decreases to $-90°$, and the magnitude decreases monotonically to zero. Thus the Polar Plot may be sketched as shown in Fig. 11-51. The dashed line represents the mirror image of the plot for $0 < \omega < \infty$ (Section 11.6, Property 2). hence it is the Polar Plot for $-\infty < \omega < 0$.

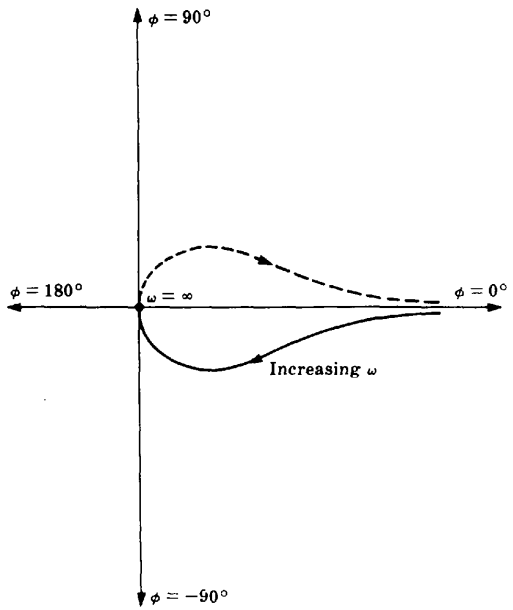


Fig. 11-51

THE NYQUIST PATH

11.39. Prove that the infinite semicircle, portion $\overline{def}$ of the Nyquist Path, maps into the origin $P(s) = 0$ in the $P(s)$-plane for all transfer functions of the form:

$$P(s) = \frac{K}{\prod_{i=1}^{n}(s + p_i)}$$

where $n > 0$, K is a constant, and $-p_i$ is any finite pole.

For $n > 0$,

$$\left|\lim_{R\to\infty} P(Re^{j\theta})\right| \equiv |P(\infty)| = \lim_{R\to\infty}\left|\frac{K}{\prod_{i=1}^{n}(Re^{j\theta} + p_i)}\right|$$

$$= \lim_{R\to\infty}\frac{|K|}{\prod_{i=1}^{n}|Re^{j\theta} + p_i|} \le \lim_{R\to\infty}\frac{|K|}{\prod_{i=1}^{n}|R - |p_i||} = 0$$

Since $|P(\infty)| \le 0$, then clearly $|P(\infty)| \equiv 0$.

11.40. Prove that the infinite semicircle, portion $\overline{def}$ of the Nyquist Path, maps into the origin $P(s) = 0$ in the $P(s)$-plane for all transfer functions of the form:

$$P(s) = \frac{K\prod_{i=1}^{m}(s + z_i)}{\prod_{i=1}^{n}(s + p_i)}$$

where $m < n$, K is a constant, and $-p_i$ and $-z_i$ are finite poles and zeros, respectively.

For $m < n$,

$$\left|\lim_{R\to\infty} P(Re^{j\theta})\right| \equiv |P(\infty)| = \lim_{R\to\infty}\left|\frac{K\prod_{i=1}^{m}(Re^{j\theta} + z_i)}{\prod_{i=1}^{n}(Re^{j\theta} + p_i)}\right|$$

$$= \lim_{R\to\infty}\frac{|K|\prod_{i=1}^{m}|Re^{j\theta} + z_i|}{\prod_{i=1}^{n}|Re^{j\theta} + p_i|} \le \lim_{R\to\infty}\frac{|K|\prod_{i=1}^{m}|R + |z_i||}{\prod_{i=1}^{n}|R - |p_i||} = 0$$

Since $|P(\infty)| \le 0$, then $|P(\infty)| \equiv 0$.

NYQUIST STABILITY PLOTS

11.41. Prove that a continuous type l system includes l infinite semicircles in the locus of its Nyquist Stability Plot. That is, show that portion $\overline{ija}$ of the Nyquist Path maps into an arc of $180l$ degrees at infinity in the $P(s)$-plane.

The transfer function of a continuous type l system has the form:

$$P(s) = \frac{B_1(s)}{s^l B_2(s)}$$

where $B_1(0)$ and $B_2(0)$ are finite and nonzero. If we let $B_1(s)/B_2(s) \equiv F(s)$, then

$$P(s) = \frac{F(s)}{s^l}$$

where $F(0)$ is finite and nonzero. Now put $s = \rho e^{j\theta}$, as required by Equation (*11.12*). Clearly, $\lim_{\rho \to 0} F(\rho e^{j\theta}) = F(0)$. Then $P(\rho e^{j\theta}) = F(\rho e^{j\theta})/\rho^l e^{jl\theta}$ and

$$\lim_{\rho \to 0} P(\rho e^{j\theta}) = \infty \cdot e^{-jl\theta} \qquad -90° \le \theta \le +90°$$

At $\theta = -90°$, the limit is $\infty \cdot e^{jl90^\circ}$. At $\theta = +90°$, the limit is $\infty \cdot e^{-jl90^\circ}$. Hence the angle subtended in the $P(s)$-plane, by mapping the locus of the infinitesimal semicircle of the Nyquist Path in the neighborhood of the origin in the s-plane, is $90l - (-90l) = 180l$ degrees, which represents l infinite semicircles in the $P(s)$-plane.

11.42. Sketch the Nyquist Stability Plot for the open-loop transfer function given by

$$GH(s) = \frac{1}{(s + p_1)(s + p_2)} \qquad p_1, p_2 > 0$$

The Nyquist Path for this type 0 system is shown in Fig. 11-52.

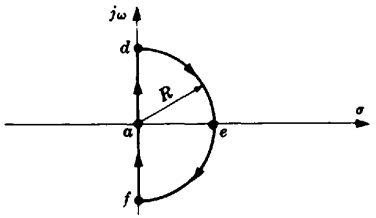


Fig. 11-52

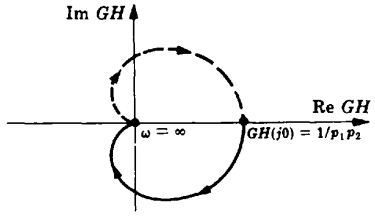


Fig. 11-53

Since there are no poles on the $j\omega$-axis, Step 2 of Section 11.8 indicates that the Polar Plot of $GH(j\omega)$ yields the image of path $\overline{ad}$ (and hence $\overline{fad}$) in the $GH(s)$-plane. Letting $s = j\omega$ for $0 < \omega < \infty$, we get

$$GH(j\omega) = \frac{1}{(j\omega + p_1)(j\omega + p_2)} = \frac{1}{\sqrt{(\omega^2 + p_1^2)(\omega^2 + p_2^2)}} \angle -\tan^{-1}\left(\frac{\omega}{p_1}\right) - \tan^{-1}\left(\frac{\omega}{p_2}\right)$$

$$GH(j0) = \frac{1}{p_1 p_2} \angle 0° \qquad \lim_{\omega \to \infty} GH(j\omega) = 0 \angle 180°$$

For $0 < \omega < \infty$, the Polar Plot passes through the third and fourth quadrants because $\phi = -[\tan^{-1}(\omega/p_1) + \tan^{-1}(\omega/p_2)]$ varies from 0° to 180° when ω increases.

From Problem 11.39, path $\overline{def}$ plots into the origin $P(s) = 0$. Therefore the Nyquist Stability Plot is a replica of the Polar Plot. This is easily sketched from the above derivations, and is shown in Fig. 11-53.

11.43. Sketch the Nyquist Stability Plot for $GH(s) = 1/s$.

The Nyquist Path for this simple type 1 system is shown in Fig. 11-54.

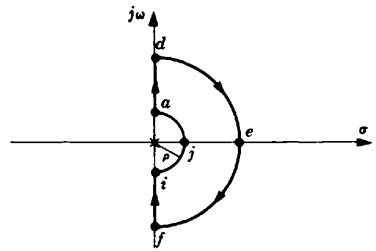


Fig. 11-54

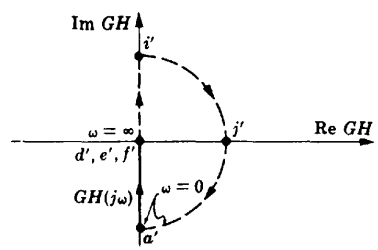


Fig. 11-55

For path $\overline{ad}$, $s = j\omega$, $0 < \omega < \infty$, and

$$GH(j\omega) = \frac{1}{j\omega} = \frac{1}{\omega} \angle -90° \qquad \lim_{\omega \to 0} GH(j\omega) = \infty \angle -90° \qquad \lim_{\omega \to \infty} GH(j\omega) = 0 \angle -90°$$

Path $\overline{def}$ maps into the origin (see Problem 11.39).

Path $\overline{f'i'}$ is the mirror image of $\overline{a'd'}$ about the real axis.

The image of path $\overline{ija}$ is determined from Equation (*11.12*), by letting $s = \lim_{\rho \to 0} \rho e^{j\theta}$, where $-90° \le \theta \le 90°$:

$$\lim_{\rho \to 0} GH(\rho e^{j\theta}) = \lim_{\rho \to 0} \left[\frac{1}{\rho} e^{-j\theta}\right] = \infty \cdot e^{-j\theta} = \infty \angle -\theta$$

For point i, $\theta = -90°$; then i maps into i' at $\infty \angle 90°$. At point j, $\theta = 0°$; then j maps into j' at $\infty \angle 0°$. Similarly, a maps into a' at $\infty \angle -90°$. Path $\overline{i'j'a'}$ could also have been obtained from the conformal mapping property of the transformation as explained in Example 11.9 plus the statement proved in Problem 11.41.

The resulting Nyquist Stability Plot is shown in Fig. 11-55.

11.44. Sketch the Nyquist Stability Plot for $GH(s) = 1/s(s + p_1)(s + p_2)$, $p_1, p_2 > 0$.

The Nyquist Path for this type 1 system is the same as that for the preceding problem. For path $\overline{ad}$, $s = j\omega$, $0 < \omega < \infty$, and

$$GH(j\omega) = \frac{1}{j\omega(j\omega + p_1)(j\omega + p_2)} = \frac{1}{\omega\sqrt{(\omega^2 + p_1^2)(\omega^2 + p_2^2)}} \angle -90° - \tan^{-1}\left(\frac{\omega}{p_1}\right) - \tan^{-1}\left(\frac{\omega}{p_2}\right)$$

$$\lim_{\omega \to 0} GH(j\omega) = \infty \angle -90° \qquad \lim_{\omega \to \infty} GH(j\omega) = 0 \angle -270° = 0 \angle +90°$$

Since the phase angle changes sign as ω increases, the plot crosses the real axis. At intermediate values of frequency, the phase angle ϕ is within the range $-90° < \phi < -270°$. Hence the plot is in the second and third quadrants. An asymptote of $GH(j\omega)$ for $\omega \to 0$ is found by writing $GH(j\omega)$ as a real plus an imaginary part, and *then* taking the limit as $\omega \to 0$:

$$GH(j\omega) = \frac{-(p_1 + p_2)}{(\omega^2 + p_1^2)(\omega^2 + p_2^2)} - \frac{j(p_1 p_2 - \omega^2)}{\omega(\omega^2 + p_1^2)(\omega^2 + p_2^2)} \qquad \lim_{\omega \to 0} GH(j\omega) = \frac{-(p_1 + p_2)}{p_1^2 p_2^2} - j\infty$$

Hence the line $GH = -(p_1 + p_2)/p_1^2 p_2^2$ is an asymptote of the Polar Plot.

Path $\overline{def}$ maps into the origin (see Problem 11.39). Path $\overline{f'i'}$ is the mirror image of $\overline{a'd'}$ about the real axis. Path $\overline{i'j'a'}$ is most easily determined by the conformal mapping property and the fact that a type 1 system has *one* infinite semicircle in its path (Problem 11.41). The resulting Nyquist Stability Plot is shown in Fig. 11-56.

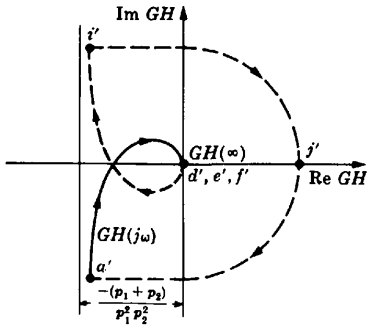


Fig. 11-56

11.45. Sketch the Nyquist Stability Plot for $GH(s) = 1/s^2$.

The Nyquist Path for this type 2 system is the same as that for the preceding problem, except there are two poles at the origin instead of one. For $\overline{ad}$,

$$GH(j\omega) = \frac{1}{j^2\omega^2} = \frac{1}{\omega^2}\angle 180° \qquad \lim_{\omega\to 0} GH(j\omega) = \infty\angle 180° \qquad \lim_{\omega\to\infty} GH(j\omega) = 0\angle 180°$$

The Polar Plot clearly lies along the negative real axis, increasing from $-\infty$ to 0 as ω increases. Path $\overline{def}$ maps into the origin and path $\overline{ija}$ maps into *two* infinite semicircles at infinity (see Problem 11.41). Since the Nyquist Path makes right turns at i and a, so does the Nyquist Stability Plot at i' and a'. The resulting locus is shown in Fig. 11-57.

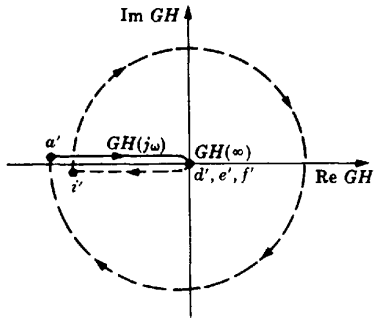


Fig. 11-57

11.46. Sketch the Nyquist Stability Plot for $GH(s) = 1/s^2(s+p)$, $p > 0$.

The Nyquist Path for this type 2 system is the same as that for the previous problem. For $\overline{ad}$,

$$GH(j\omega) = \frac{1}{j^2\omega^2(j\omega+p)} = \frac{1}{\omega^2\sqrt{\omega^2+p^2}}\angle -180° - \tan^{-1}\left(\frac{\omega}{p}\right)$$

$$\lim_{\omega\to 0} GH(j\omega) = \infty\angle -180° \qquad \lim_{\omega\to\infty} GH(j\omega) = 0\angle -270°$$

For $0 < \omega < \infty$ the phase angle varies continuously from $-180°$ to $-270°$; thus the plot lies in the second quadrant. The remainder of the Nyquist Path is mapped into the GH-plane as in the preceding problem. The resulting Nyquist Stability Plot is shown in Fig. 11-58.

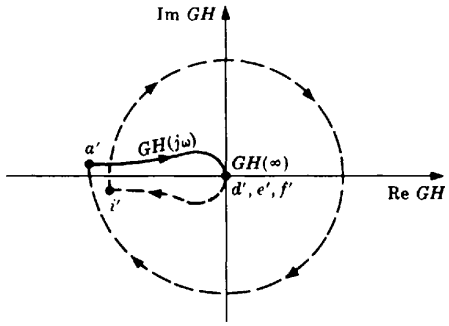


Fig. 11-58

11.47. Sketch the Nyquist Stability Plot for $GH(s) = 1/s^4(s+p)$, $p > 0$.

There are four poles at the origin in the s-plane, and the Nyquist Path is the same as that of the previous problem. The Polar Plot for this system was determined in Problem 11.38. The remainder of the Nyquist Path is mapped using the results of Problems 11.39 and 11.41, and the conformal mapping property. The resulting Nyquist Stability Plot is given in Fig. 11-59.

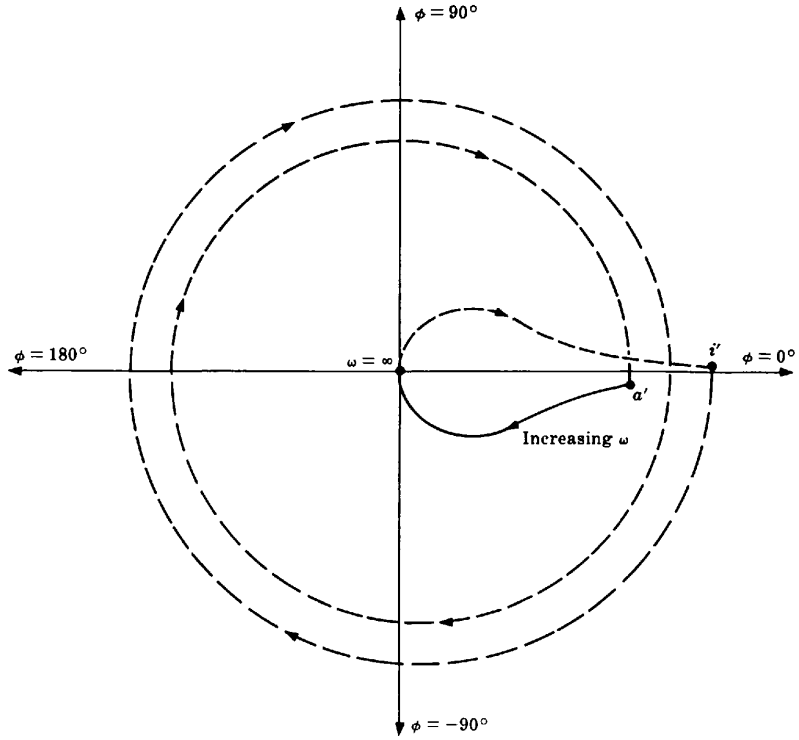


Fig. 11-59

11.48. Sketch the Nyquist Stability Plot for $GH(s) = e^{-Ts}/(s+p)$, $p > 0$.

The e^{-Ts} term represents a time delay of T seconds in the forward or feedback path. For example, a signal flow graph of such a system can be represented as in Fig. 11-60.

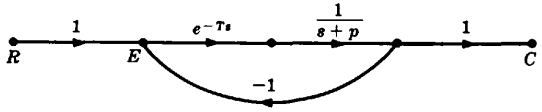


Fig. 11-60

The Nyquist Stability Plot for $1/(s+1)$ was drawn in Example 11.8. The plot is modified by inclusion of the e^{-T_s} term in the following manner. For path $\overline{ad}$,

$$GH(j\omega) = \frac{e^{-T_{j\omega}}}{j\omega + p} = \frac{1}{\sqrt{\omega^2 + p^2}} \angle -\tan^{-1}\left(\frac{\omega}{p}\right) - T\omega \qquad GH(j0) = \frac{1}{p}\angle 0°$$

The limit of $GH(j\omega)$ as $\omega \to \infty$ does not exist. But $\lim_{\omega\to\infty}|GH(j\omega)| = 0$ and $|GH(j\omega)|$ decreases monotonically as ω increases. The phase angle term

$$\phi(\omega) = -\tan^{-1}\left(\frac{\omega}{p}\right) - T\omega$$

revolves repeatedly about the origin between 0° and $-360°$ as ω increases. Therefore the Polar Plot is a decreasing spiral, beginning at $(1/p)\angle 0°$ and approaching the origin in a CW direction. The points where the locus crosses the negative real axis are determined by letting $\phi = -180° = -\pi$ radians:

$$-\pi = -\tan^{-1}\left(\frac{\omega_\pi}{p}\right) - T\omega_\pi$$

or $\omega_\pi = p\tan(T\omega_\pi)$, which is easily solved when p and T are known. The remainder of the Nyquist Path is mapped using the results of Problems 11.41 and 11.42. The Nyquist Stability Plot is shown in Fig. 11-61. The image of path $\overline{fa}$ ($s = -j\omega$) has been omitted for clarity.

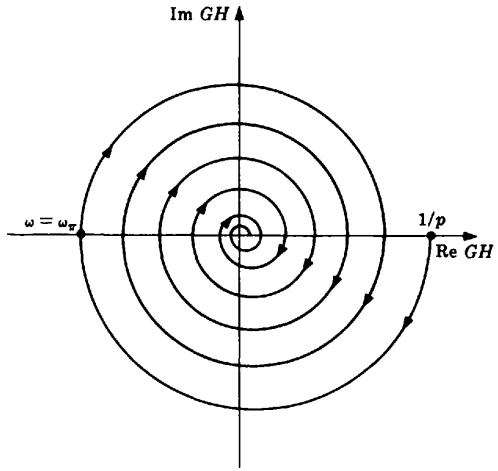


Fig. 11-61

11.49. Sketch the Nyquist Stability Plot for $GH(s) = 1/(s^2 + a^2)$.

The poles of $GH(s)$ are at $s = \pm ja \equiv \pm j\omega_0$. The Nyquist Path for this system is therefore as shown in Fig. 11-62.

For path $\overline{ab}$, $\omega < a$ and

$$GH(j\omega) = \frac{1}{a^2 - \omega^2}\angle 0° \qquad GH(j0) = \frac{1}{a^2}\angle 0° \qquad \lim_{\omega\to a} GH(j\omega) = \infty\angle 0°$$

For path $\overline{bc}$, let $s \equiv ja + \rho e^{j\theta}$, $-90° \le \theta \le 90°$; then

$$\lim_{\rho\to 0} GH(ja + \rho e^{j\theta}) = \lim_{\rho\to 0}\left[\frac{1}{\rho e^{j\theta}(2ja + \rho e^{j\theta})}\right] = -j\infty \cdot e^{-j\theta} = \infty\angle -\theta^0 - 90°$$

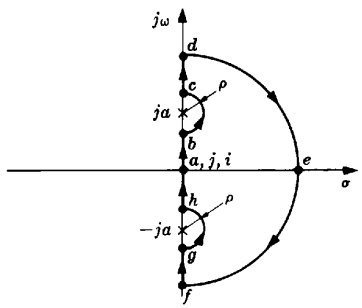


Fig. 11-62

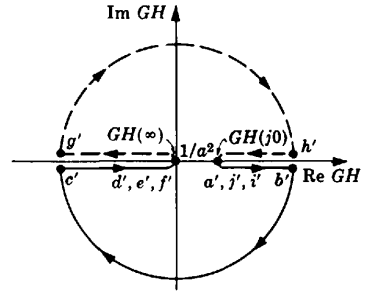


Fig. 11-63

At $\theta = -90°$ the limit is $\infty\angle 0°$; at $\theta = 0°$ it is $\infty\angle -90°$; at $\theta = 90°$ it is $\infty\angle -180°$.

For path $\overline{cd}$, $\omega > a$ and

$$\lim_{\omega\to a} GH(j\omega) = \infty\angle 180° \qquad \lim_{\omega\to\infty} GH(j\omega) = 0\angle 180°$$

Path $\overline{def}$ maps into the origin by Problem 11.39, and $\overline{f'g'h'a'}$ is the mirror image of $\overline{a'b'c'd'}$ about the real axis. The resulting Nyquist Stability Plot is shown in Fig. 11-63.

11.50. Sketch the Nyquist Stability Plot for $GH(s) = (s - z_1)/s(s + p)$, $z_1, p > 0$.

The Nyquist Path for this type 1 system is the same as that for Problem 11.43. For path $\overline{ad}$,

$$GH(j\omega) = \frac{j\omega - z_1}{j\omega(j\omega + p)} = \frac{\sqrt{\omega^2 + z_1^2}}{\omega\sqrt{\omega^2 + p^2}} \angle 90° - \tan^{-1}\left[\frac{\omega(p + z_1)}{pz_1 - \omega^2}\right]$$

where we have used

$$\tan^{-1}x \pm \tan^{-1}y \equiv \tan^{-1}\left[\frac{x \pm y}{1 \mp xy}\right]$$

Now

$$\lim_{\omega\to 0} GH(j\omega) = \infty\angle +90° \qquad GH(j\sqrt{pz_1}) = \frac{1}{p\sqrt{pz_1}}\angle 0° \qquad \lim_{\omega\to\infty} GH(j\omega) = 0\angle -90°$$

Thus the locus comes down in the first quadrant, crosses the positive real axis into the fourth quadrant, and approaches the origin from an angle of $-90°$.

Path $\overline{def}$ maps into the origin, and $\overline{ija}$ maps into one semicircle at infinity. The resulting plot is shown in Fig. 11-64.

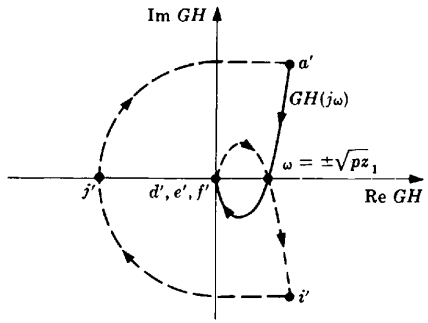


Fig. 11-64

NYQUIST STABILITY CRITERION

11.51. Prove the Nyquist Stability Criterion.

Equation (*11.1*) states that the number of CW encirclements N_0 of the origin made by a closed P contour in the P-plane, mapped from a closed complex-plane contour, is equal to the number of zeros Z_0 minus the number of poles P_0 of P enclosed by the complex-plane contour: $N_0 = Z_0 - P_0$. This has been proven in Problem 11.26.

Now let $P \equiv 1 + GH$. Then the origin for $1 + GH$ in the GH-plane is at $GH = -1$. (See Example 11.6 and Problem 11.33.) Hence let N denote the number of CW encirclements of this $-1 + j0 \equiv (-1,0)$ point, and let the complex-plane contour be the Nyquist Path defined in Section 11.7. Then $N = Z_0 - P_0$, where Z_0 and P_0 are the number of zeros and poles of $1 + GH$ enclosed by the Nyquist Path. P_0 is also the number of poles of GH enclosed, since if $GH \equiv N/D$, then $1 + GH = 1 + N/D = (D + N)/D$. That is, GH and $1 + GH$ have the same denominator.

We know from Chapter 5 that a continuous (or discrete) feedback system is absolutely stable if and only if the zeros of the characteristic polynomial $1 + GH$ (the roots of the characteristic equation $1 + GH = 0$) are in the LHP (or unit circle), that is, $Z_0 = 0$. Therefore $N = -P_0$, and clearly $P_0 \geq 0$.

11.52. Extend the Nyquist Stability Criterion to a larger class of continuous linear systems than those already considered in this chapter.

The Nyquist Stability Criterion has been extended by Desoer [5]. The following statement is a modification of this generalization, found with its proof in the reference.

A Generalized Nyquist Stability Criterion: Consider the linear time-invariant system described by the block diagram in Fig. 11-65. If $g(t)$ satisfies the conditions given below and the Nyquist Stability Plot of $G(s)$ *does not* enclose the $(-1,0)$ point, then the system is *stable*. If the $(-1,0)$ point is enclosed, the system is unstable.

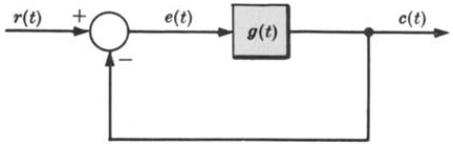


Fig. 11-65

1. $G(s)$ represents a causal, linear time-invariant system element.
2. The input-output relationship for $g(t)$ is

$$c(t) = c_a(t) + \int_0^t g(t-\tau)e(\tau)\,d\tau \qquad t \geq 0$$

where $c_a(t)$, the free response of the system $g(t)$, is bounded for all $t \geq 0$ and all initial conditions, and approaches a finite value dependent upon the initial conditions as $t \to \infty$.

3. The unit impulse response $g(t)$ is

$$g(t) = [k + g_1(t)]\mathbf{1}(t)$$

where $k \geq 0$, $\mathbf{1}(t)$ is the unit step function, $g_1(t)$ is bounded and integrable for all $t \geq 0$, and $g_1(t) \to 0$ as $t \to \infty$.

These conditions are fulfilled very often by physical systems described by ordinary and partial differential equations, and differential-difference equations. The form of the closed-loop block diagram given in Fig. 11-65 is not necessarily restrictive. Many systems of interest can be transformed into this configuration.

11.53. Suppose the Nyquist Path for $GH(s) = 1/s(s + p)$ were modified so that the pole at the origin is enclosed as shown in Fig. 11-66. How does this modify application of the Nyquist Stability Criterion?

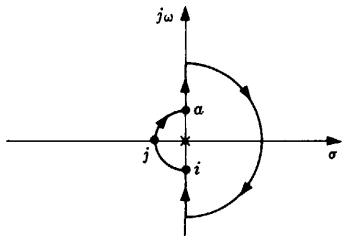


Fig. 11-66

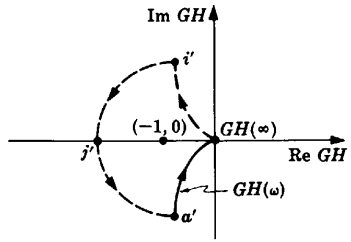


Fig. 11-67

The Polar Plot remains the same, but the image of path $\overline{ija}$ makes *left* instead of right turns at i' and a', just as in the Nyquist Path. The Nyquist Stability Plot is therefore given by Fig. 11-67. Clearly, $N = -1$. But since the pole of GH at the origin is enclosed by the Nyquist Path, then $P_0 = 1$, and $Z_0 = N + P_0 = -1 + 1 = 0$. Therefore the system is stable. Application of the Nyquist Stability Criterion does not depend on the path chosen in the s-plane.

11.54. Is the system of Problem 11.42 stable or unstable?

Shading the region to the right of the contour in the prescribed direction yields Fig. 11-68. It is clear that $N = 0$. The $(-1,0)$ point is not in the shaded region. Now, since $p_1 > 0$ and $p_2 > 0$, then $P_0 = 0$. Therefore $N = -P_0 = 0$, or $Z_0 = N + P_0 = 0$, and the system is stable.

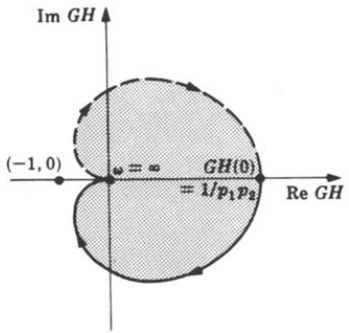


Fig. 11-68

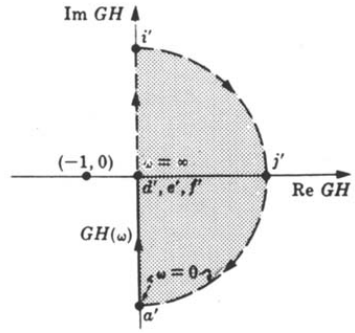


Fig. 11-69

11.55. Is the system of Problem 11.43 stable or unstable?

The region to the right of the contour has been shaded in Fig. 11-69. The $(-1,0)$ point is not enclosed, and $N = 0$. Since $P_0 = 0$, then $Z_0 = P_0 + N = 0$, and the system is stable.

11.56. Determine the stability of the system of Problem 11.44.

The region to the right of the contour has been shaded in Fig. 11-70. If the $(-1,0)$ point lies to the left of point k, then $N=0$; if it lies to the right, then $N=1$. Since $P_0=0$, then $Z_0=0$ or 1. Hence the system is stable if and only if the $(-1,0)$ point lies to the left of point k. Point k can be determined by solving for $GH(j\omega_\pi)$, where

$$-\pi = \frac{-\pi}{2} - \tan^{-1}\left(\frac{\omega_\pi}{p_1}\right) - \tan^{-1}\left(\frac{\omega_\pi}{p_2}\right)$$

ω_π is easily determined from this equation when p_1 and p_2 are given.

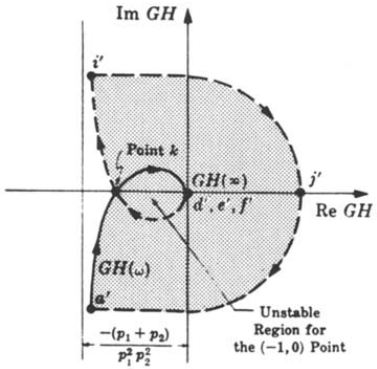


Fig. 11-70

11.57. Determine the stability of the system of Problem 11.46.

The region to the right of the contour has been shaded in Fig. 11-71. Clearly, $N=1$, $P_0=0$, and $Z_0=1+0=1$. Hence the system is unstable for all $p>0$.

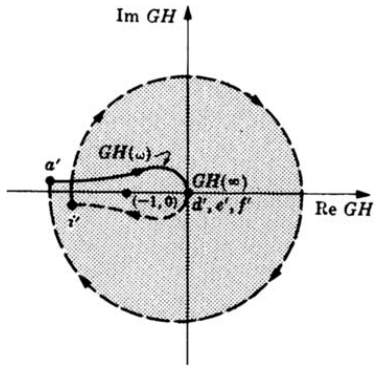


Fig. 11-71

11.58. Determine the stability of the system of Problem 11.47.

The region to the right of the contour has been shaded in Fig. 11-72.
It is clear that $N>0$. Since $P_0=0$ for $p>0$, then $N\neq -P_0$. Hence the system is unstable.

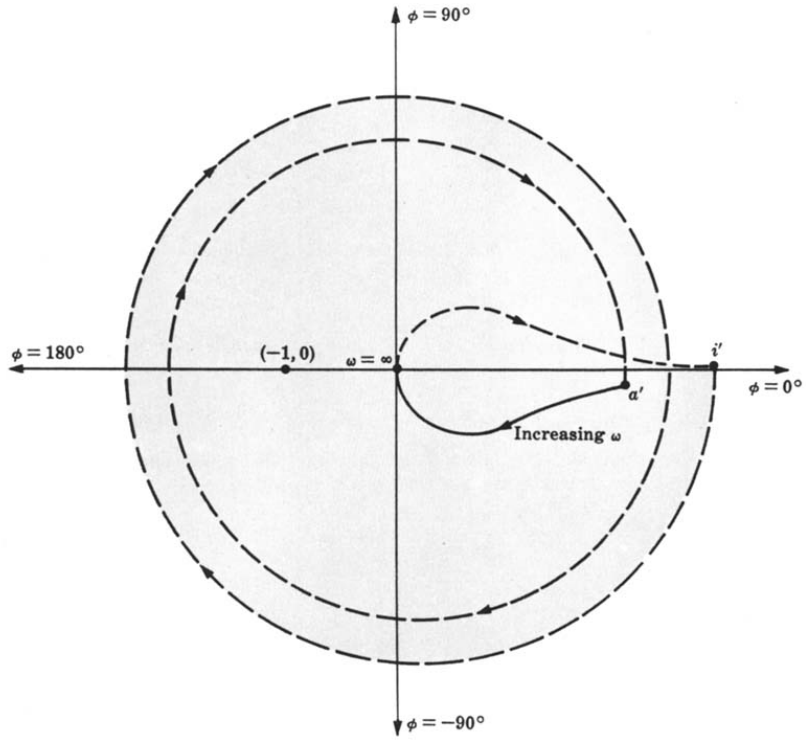


Fig. 11-72

RELATIVE STABILITY

11.59. Determine: (a) the phase crossover frequency ω_π, (b) the gain crossover frequency ω_1, (c) the gain margin, and (d) the phase margin for the system of Problem 11.44 with $p_1=1$ and $p_2=\frac{1}{2}$.

Mathcad

(a) Letting $\omega=\omega_\pi$, we have

$$\phi(\omega_\pi) = -\pi = \frac{-\pi}{2} - \tan^{-1}\omega_\pi - \tan^{-1}2\omega_\pi = \frac{-\pi}{2} - \tan^{-1}\left(\frac{3\omega_\pi}{1-2\omega_\pi^2}\right)$$

or $3\omega_\pi/(1-2\omega_\pi^2)=\tan(\pi/2)=\infty$. Hence $\omega_\pi=\sqrt{\frac{1}{2}}=0.707$.

(b) From $|GH(\omega_1)|=1$, we have $1/\omega_1\sqrt{(\omega_1^2+1)(\omega_1^2+0.25)}=1$ or $\omega_1=0.82$.

(c) The gain margin $1/|GH(\omega_\pi)|$ is easily determined from the graph, as shown in Fig. 11-73. It can also be calculated analytically: $|GH(\omega_\pi)|=|GH(j0.707)|=4/3$; hence gain margin $=3/4$.

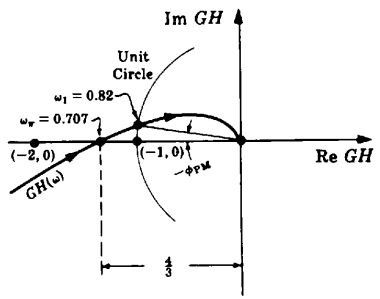


Fig. 11-73

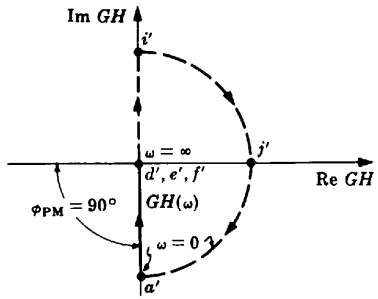


Fig. 11-74

(*d*) The phase margin is easily determined from the graph, or calculated analytically:

$$\arg GH(\omega_1) = \arg GH(0.82) = -90° - \tan^{-1}(0.82) - \tan^{-1}(1.64) = -187.8°$$

Hence $\phi_{PM} = 180° + \arg GH(\omega_1) = -7.8°$. Negative phase margin means that the system is unstable.

11.60. Determine the gain and phase margins for the system of Problem 11.43 ($GH = 1/s$).

The Nyquist Stability Plot of $1/s$ never crosses the negative real axis as shown in Fig. 11-74; hence the gain margin is undefined for this system. The phase margin is $\phi_{PM} = 90°$.

M- AND *N*-CIRCLES

11.61. Prove Equations (*11.18*) and (*11.19*), which give the radius and center of an *M*-circle, respectively.

Let $G(\omega) \equiv x + jy$. Then

$$M \equiv \left|\frac{G(\omega)}{1+G(\omega)}\right| = \left|\frac{x+jy}{1+x+jy}\right|$$

Squaring both sides and rearranging yields

$$\left[x - \left(\frac{M^2}{1-M^2}\right)\right]^2 + y^2 = \left(\frac{M}{1-M^2}\right)^2 \qquad M < 1$$

$$\left[x + \left(\frac{M^2}{M^2-1}\right)\right]^2 + y^2 = \left(\frac{M}{M^2-1}\right)^2 \qquad M > 1$$

For M = constant, these are equations of circles with radii $|M/(M^2-1)|$ and centers at $(-M^2/(M^2-1), 0)$.

11.62. Prove Equation (*11.20*).

The transfer function G for the second-order continuous system whose signal flow graph is shown in Fig. 11-75 is $G = \omega_n^2/s(s+2\zeta\omega_n)$. Now

$$M^2 = \left|\frac{G}{1+G}\right|^2 = \frac{\omega_n^4}{(\omega_n^2-\omega^2)^2 + 4\zeta^2\omega_n^2\omega^2}$$

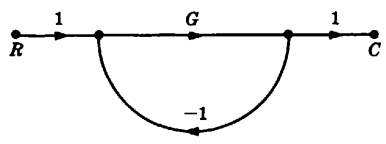


Fig. 11-75

To find ω_p, we maximize the above expression:

$$\frac{d}{d\omega}(M^2) = \frac{\omega_n^4\left[2(\omega_n^2-\omega^2)(-2\omega) + 8\zeta^2\omega_n^2\omega\right]}{\left[(\omega_n^2-\omega^2)^2 + 4\zeta^2\omega_n^2\omega^2\right]^2} = 0$$

from which $\omega = \omega_p = \pm\omega_n\sqrt{1-2\zeta^2}$. Hence for $0 \le \zeta \le 0.707$,

$$M_p = \left[\frac{\omega_n^4}{\left[\omega_n^2 - \omega_n^2(1-2\zeta^2)\right]^2 + 4\zeta^2\omega_n^4(1-2\zeta^2)}\right]^{1/2} = \frac{1}{2\zeta\sqrt{1-\zeta^2}}$$

11.63. Prove Equations (*11.21*) and (*11.22*), which give the radius and center of an *N*-circle.

Let $G(\omega) \equiv x + jy$. Then

$$\frac{C(\omega)}{R(\omega)} = \frac{x^2+x+y^2+jy}{(1+x)^2+y^2} \qquad \text{and} \qquad N \equiv \frac{\mathrm{Im}(C/R)(\omega)}{\mathrm{Re}(C/R)(\omega)} = \frac{y}{x^2+x+y^2}$$

which yields

$$\left(x+\frac{1}{2}\right)^2 + \left(y - \frac{1}{2N}\right)^2 = \frac{1}{4}\left(1+\frac{1}{N^2}\right)$$

For N equal to a constant parameter, this is the equation of a circle with radius $\sqrt{\frac{1}{4} + (1/2N)^2}$ and center at $(-\frac{1}{2}, 1/2N)$.

11.64. Find M_p and ζ for the unity feedback system given by $G = 1/s(s+1)$.

The general open-loop transfer function for the second-order system is $G = \omega_n^2/s(s+2\zeta\omega_n)$. Then $\omega_n = 1$, $\zeta = 0.5$, and $M_p = 1/(2\zeta\sqrt{1-\zeta^2}) = 0.866$.

MISCELLANEOUS PROBLEMS

11.65. Determine the Polar Plot for

$$P(z) = \frac{z}{z-1}$$

for a sampling period $T = 1$.

The solution requires mapping the strip from $-j\omega_s/2$ to $j\omega_s/2$ on the $j\omega$-axis of the s-plane or, equivalently, $\omega = -\pi$ to $\omega = \pi$ radians on the unit circle of the z-plane, into the $P(e^{j\omega})$-plane. We have $P(e^{\pm j\pi}) = 0.5\angle 0°$, and $P(e^{j0}) = \infty\angle \pm 90°$. Evaluation of $P(e^{j\omega})$ for several values of ω between $-\pi$ and 0 results in a straight line parallel to the imaginary axis in the P-plane, as shown in Fig. 11-76, where the segments a to b and g to a map the corresponding segments of the unit circle in Fig. 11-13.

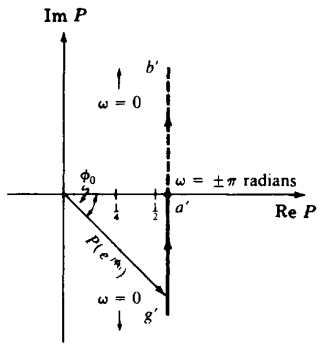


Fig. 11-76

11.66. Determine the Polar Plot of the type 0 discrete-time system open-loop transfer function

$$GH(z) = \frac{\frac{3}{8}(z+1)(z+\frac{1}{3})K}{z(z+\frac{1}{2})}$$

for $K = 1$ and $T = 1$.

In this case, the Polar Plot has been drawn by computer, as illustrated in Fig. 11-77. The computer program evaluates $GH(e^{j\omega})$ for values of $\omega T = \omega$ in the range $-\pi$ to π radians, separates each result into real and imaginary parts (Complex Form), and then generates the rectangular plot from these coordinates.

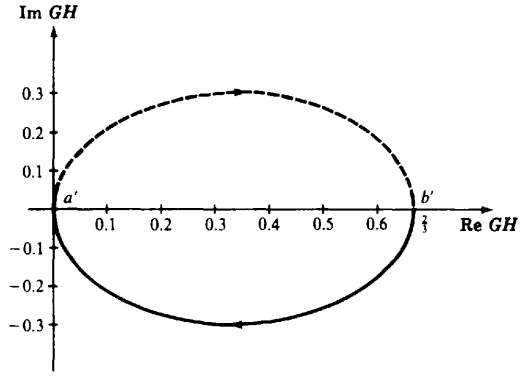


Fig. 11-77

11.67. Determine the Polar Plot of the type 1 discrete-time system open-loop transfer function

$$GH(z) = \frac{K(z+1)^2}{(z-1)(z+\frac{1}{3})(z+\frac{1}{2})}$$

for $K = 1$ and $T = 1$.

As in Problem 11.66, the Polar Plot given in Fig. 11-78 was generated by computer, exactly in the same manner as described in the previous problem.

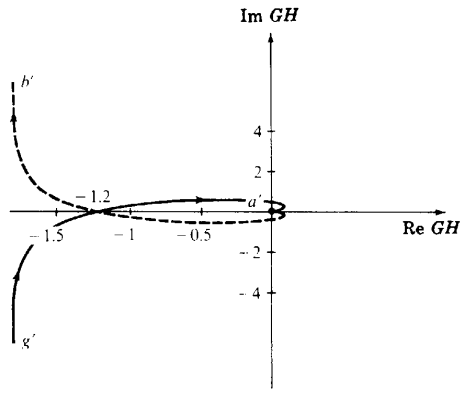


Fig. 11-78

11.68. Determine the absolute stability of the system given in Examples 11.11 and 11.14, for $K \geq 2$ and $T = 1$.

The Nyquist Stability Plot for $K = 2$ is given in Fig. 11-79. The region to the right has been shaded and the plot goes directly through $(-1, 0)$. Thus $N > 0$ and $N \neq -P_0$, which is zero for this problem. Therefore the system is marginally stable for $K = 2$. For $K > 2$, the $(-1, 0)$ point is completely enclosed, $N = 1$, and the closed-loop system is unstable.

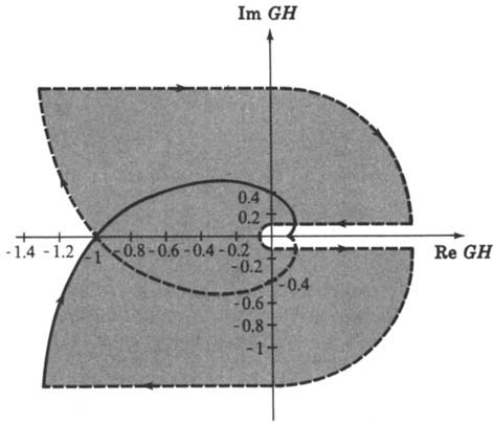


Fig. 11-79

11.69. Determine the Nyquist Stability Plot of the system given in Problem 11.65.

We note that $P(z) = z/(z-1)$ has a pole at 1, so we must begin by mapping the segment b to c of the infinitesimal semicircle near $z = 1$ in Fig. 11-13 into the P-plane. Overall, we have a conformal mapping, so the plot must turn right at b'. Between b and c, $z = 1 + \rho e^{j\phi}$, with ϕ increasing from $-90°$ to $0°$. Therefore

$$P(1+\rho e^{j\phi}) = \frac{1+\rho e^{j\phi}}{\rho e^{j\phi}} \qquad \text{and} \qquad \lim_{\rho\to 0} P(1+\rho e^{j\phi}) = \frac{1}{\lim_{\rho\to 0}\rho e^{j\phi}} = \infty\angle -\phi$$

Therefore the arc from b to c in the z-plane maps into the infinite semicircle b' to c', from $+90°$ back to $0°$, shown in Fig. 11-80. To obtain the mapping of the line from c to d in Fig. 11-13, we note that this is the mapping of $P(z)$ from $z = 1\angle 0°$ to $z = \infty\angle 0°$ (constant angle ϕ), that is,

$$P(1) = \infty\angle 0° \qquad \text{to} \qquad \lim_{\alpha\to\infty}\left(\frac{1+\alpha}{1+\alpha-1}\right) = \lim_{\alpha\to\infty}\left(\frac{1+\alpha}{\alpha}\right) = 1\angle 0°$$

where we have replaced z in $P(z)$ by $1 + \alpha$, in obtaining the limit. The resulting mapping is shown as the line from c' to d' ($\infty \to 1$) in Fig. 11-80.

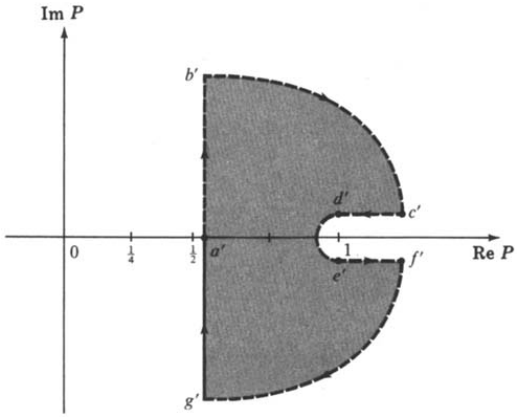


Fig. 11-80

The infinite circle from $0°$ to $-360°$, from d to e in Fig. 11-13, maps into an infinitesimal semicircle around the point $z = 1$ in the P-plane, because

$$P(Re^{j\phi}) = \frac{Re^{j\phi}}{Re^{j\phi}-1} = \frac{e^{j\phi}}{e^{j\phi}-\dfrac{1}{R}}$$

and $P \to 1$ as $R \to \infty$ for any ϕ, and a few evaluations of $\arg P(Re^{j\phi})$ at values of ϕ between $0°$ and $-360°$ show that the limit is approached from values in the first quadrant of P when $0 < \phi < -180°$, and the fourth quadrant when $-180° < \phi < -360°$, with $P(Re^{j\phi}) = 1/(1 + 1/R) < 1$ for $R > 0$ at $\phi = 180°$. The resulting arc is shown as d' to e' in Fig. 11-80.

Arc e' to f' in Fig. 11-80 is obtained in the same manner as that for c' to d', taking the limits of $(\alpha + 1)/\alpha$ as $\alpha \to \infty$ and 0. And the final closure of the Nyquist Stability Plot, arc f' to g', is obtained in the same manner as that from b' to c', as shown.

11.70. For $GH = P = z/(z-1)$ in Problem 11.69, is the closed-loop system stable?

The region to the right of the contour in Fig. 11-80 has been shaded and it does not enclose the $(-1,0)$ point. Therefore $N \le 0$. The only pole of GH is at $z = 1$, which is not outside the unit circle. Thus $P_0 = 0$, $N = -P_0 = 0$, and the system is absolutely stable.

11.71. Determine the stability of the system given in Problem 11.66.

The open-loop transfer function is

$$GH = \frac{\frac{3}{8}(z+1)(z+\frac{1}{3})}{z(z+\frac{1}{2})}$$

The Polar Plot of GH is given in Fig. 11-77, which is the mapping of arcs a to b and g to a of Fig. 11-13. There are no poles of GH on the unit circle, so the infinitesimal arcs b to c and f to g in Fig. 11-13 are not needed. Setting $z = 1 + \alpha$ and using the same limiting procedures illustrated in Problem 11.70, the straight lines to and from infinity, b to d and e to f in Fig. 11-13, map into the lines from b to d and e to f between $\text{Re}\, GH = \frac{3}{8}$ and $\frac{2}{3}$. Similarly, with z replaced by $1 + Re^{j\phi}$ and $R \to \infty$, the infinite arc from d to e maps into the infinitesimal semicircle about $\text{Re}\, GH = \frac{3}{8}$, all as shown in Fig. 11-81.

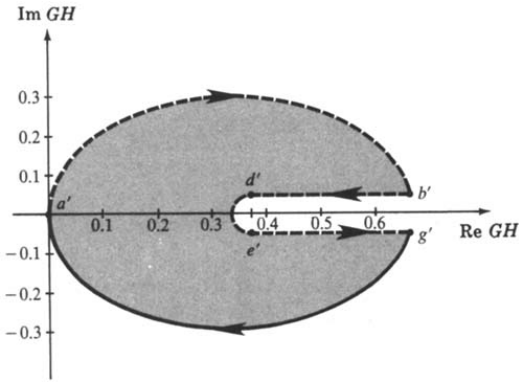


Fig. 11-81

The $(-1,0)$ point is not enclosed by this contour, as shown, $N = 0$, $P_0 = 0$, and the closed-loop system is absolutely stable.

11.72. Determine the stability of the system given in Problem 11.67.

The open-loop transfer function is

$$GH = \frac{(z+1)^2}{(z-1)(z+\frac{1}{3})(z+\frac{1}{2})}$$

The Polar Plot of GH is given in Fig. 11-78. Completion of the closed contour mapping of the exterior of the unit circle in the z-plane (Fig. 11-13) closely parallels that described in Problem 11.69 and Example 11.11. In this case, the $(-1,0)$ point is enclosed once by the contour, that is, $N = 1$. Since $P_0 = 0$ and $Z_0 = N + P_0 = 1$, then one zero of $1 + GH$ is outside the unit circle of the z-plane and the closed-loop system is therefore unstable (Fig. 11-82).

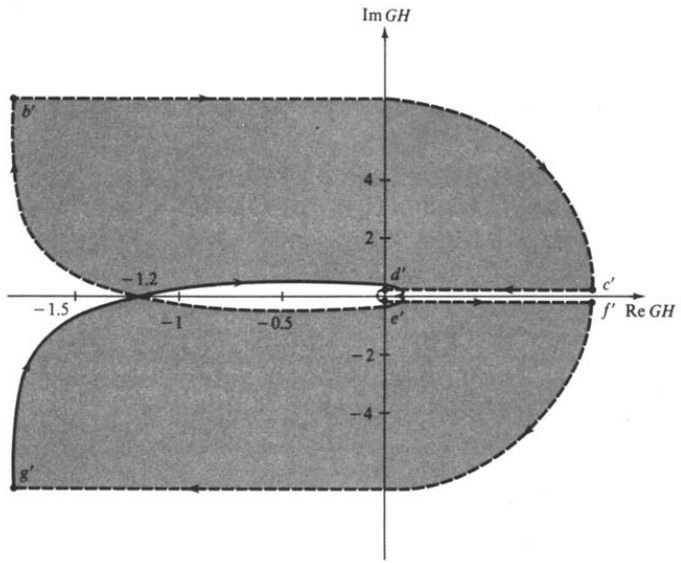


Fig. 11-82

Supplementary Problems

11.73. Let $T = 2$ and $p = 5$ in the system of Problem 11.48. Is this system stable?

11.74. Is the system of Problem 11.49 stable or unstable?

11.75. Is the system of Problem 11.50 stable or unstable?

11.76. Sketch the Polar Plot for $GH = \dfrac{K(s+z_1)(s+z_2)}{s^3(s+p_1)(s+p_2)}$, $z_i, p_i > 0$.

11.77. Sketch the Polar Plot for $GH = \dfrac{K}{(s+p_1)(s+p_2)(s+p_3)}$, $p_i > 0$.

11.78. Find the closed-loop frequency response of the unity feedback system described by $G = \dfrac{10(s+0.5)}{s^2(s+1)(s+10)}$, using M- and N-circles.

11.79. Sketch the Polar Plot for $GH = \dfrac{K(s+z_1)}{s^2(s+p_1)(s+p_2)(s+p_3)}$, $z_1, p_i > 0$.

11.80. Sketch the Nyquist Stability Plot for $GH = \dfrac{Ke^{-Ts}}{s(s+1)}$.

11.81. Sketch the Polar Plot for $GH = \dfrac{s+z_1}{s(s+p_1)}$, $z_1, p_1 > 0$.

11.82. Sketch the Polar Plot for $GH = \dfrac{s+z_1}{s(s+p_1)(s+p_2)}$, $z_1, p_i > 0$.

11.83. Sketch the Polar Plot for $GH = \dfrac{K}{s^2(s+p_1)(s+p_2)}$, $p_i > 0$.

11.84. Sketch the Polar Plot for $GH = \dfrac{s+z_1}{s^2(s+p_1)}$, $z_1, p_1 > 0$.

11.85. Sketch the Polar Plot for $GH = \dfrac{s+z_1}{s^2(s+p_1)(s+p_2)}$, $z_1, p_i > 0$.

11.86. Sketch the Polar Plot for $GH = \dfrac{(s+z_1)(s+z_2)}{s^2(s+p_1)(s+p_2)(s+p_3)}$, $z_i, p_i > 0$.

11.87. Sketch the Polar Plot for $GH = \dfrac{K}{s^3(s+p_1)(s+p_2)}$, $p_i > 0$.

11.88. Sketch the Polar Plot for $GH = \dfrac{(s+z_1)}{s^3(s+p_1)(s+p_2)}$, $z_1, p_i > 0$.

11.89. Sketch the Polar Plot for $GH = \dfrac{s+z_1}{s^4(s+p_1)}$, $z_1, p_1 > 0$.

11.90. Sketch the Polar Plot for $GH = \dfrac{e^{-Ts}(s+z_1)}{s^2(s+p_1)}$, $z_1, p_1 > 0$.

11.91. Sketch the Polar Plot for $GH = \dfrac{e^{-Ts}(s+z_1)}{s^2(s^2+a)(s^2+b)}$, $z_1, a, b > 0$.

11.92. Sketch the Polar Plot for $GH = \dfrac{(s-z_1)}{s^2(s+p_1)}$, $z_1, p_1 > 0$.

11.93. Sketch the Polar Plot for $GH = \dfrac{s}{(s+p_1)(s-p_2)}$, $p_i > 0$.

11.94. The various portions of the Nyquist Path for continuous systems are illustrated in Fig. 11-12 and the different segments are defined mathematically by Equations (*11.5*) through (*11.12*). Write the corresponding equations for each segment of the Nyquist Path for the discrete-time systems given in Fig. 11-13. (One of these was given in Example 11.11. Also see Problems 11.69 and 11.70.)

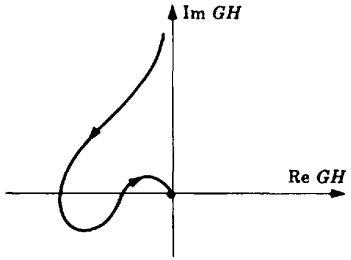
Answers to Some Supplementary Problems

11.73. Yes

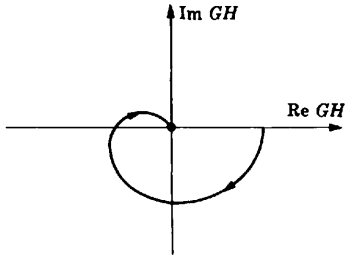
11.74. Unstable

11.75. Unstable

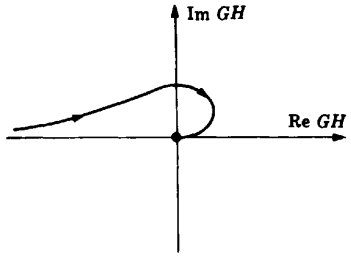
11.76.



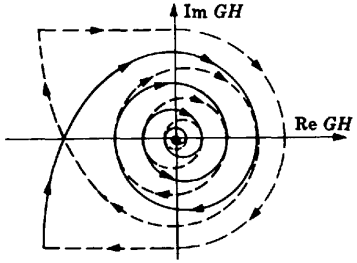
11.77.



11.79.



11.80.



Chapter 12

Nyquist Design

12.1 DESIGN PHILOSOPHY

Design by analysis in the frequency domain using Nyquist techniques is performed in the same general manner as all other design methods in this book: appropriate compensation networks are introduced in the forward or feedback paths and the behavior of the resulting system is critically analyzed and reanalyzed. In this manner, the Polar Plot is shaped and reshaped until performance specifications are met. The procedure is greatly facilitated when computer programs for generating Polar Plots are used.

Since the Polar Plot is a graph of the open-loop frequency response function $GH(\omega)$, many types of compensation components can be used in either the forward or feedback path, becoming part of either G or H. Often, compensation in only one path, or a combination of both cascade and feedback compensation, can be used to satisfy specifications. Cascade compensation is emphasized in this chapter.

12.2 GAIN FACTOR COMPENSATION

It was pointed out in Chapter 5 that an unstable feedback system can sometimes be stabilized, or a stable system destabilized, by appropriately adjusting the gain factor K of GH. The root-locus method of Chapters 13 and 14 vividly illustrates this phenomenon, but it is also evidenced in Nyquist Stability Plots.

EXAMPLE 12.1. Figure 12-1 indicates an unstable *continuous* system when the gain factor is K_1, where

$$GH(s) = \frac{K_1}{s(s+p_1)(s+p_2)} \qquad p_1, p_2, K_1 > 0 \qquad P_0 = 0 \qquad N = 2$$

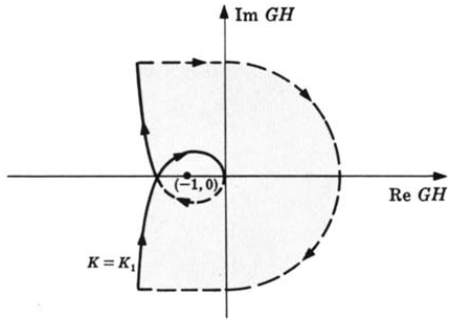


Fig. 12-1

A sufficient decrease in the gain factor to K_2 ($K_2 < K_1$) stabilizes the system, as illustrated in Fig. 12-2.

$$GH(s) = \frac{K_2}{s(s+p_1)(s+p_2)} \qquad 0 < K_2 < K_1 \qquad P_0 = 0 \qquad N = 0$$

Further decrease of K does not alter stability.

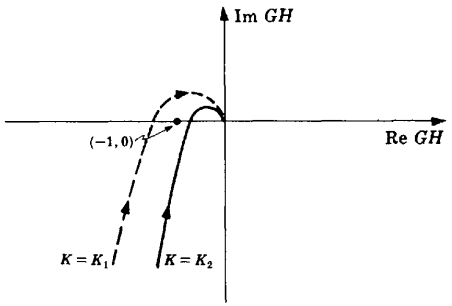


Fig. 12-2

EXAMPLE 12.2. The type 1 *discrete-time* control system with

$$GH_1 = \frac{1}{(z-1)(z-\frac{1}{2})}$$

is unstable, as shown in Fig. 11-79 and Problem 11.68. That is, the open-loop transfer function

$$GH = \frac{K/4}{(z-1)(z-\frac{1}{2})}$$

was found to be unstable for $K \geq 2$. Therefore gain factor compensation can be used to stabilize GH_1, by attenuating the gain factor $K_1 = 1$ of GH_1 by a factor less than 0.5. For example, if the attenuator is given a value of 0.25, the resulting $GH \equiv GH_2$ would have the Nyquist Stability Plot in Fig. 11-25, shown in Example 11.14 to represent a stable system.

EXAMPLE 12.3. The stable region for the $(-1,0)$ point in Fig. 12-3 is indicated by the portion of the real axis in the unshaded area:

$$GH(s) = \frac{K(s+z_1)(s+z_2)}{s^2(s+p_1)(s+p_2)(s+p_3)} \qquad z_1, z_2 > 0 \qquad p_i > 0 \qquad P_0 = 0$$

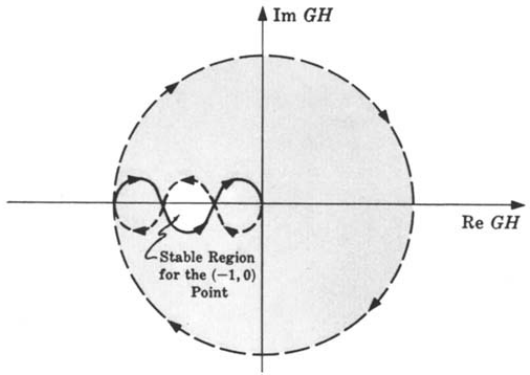


Fig. 12-3

If the $(-1,0)$ point falls in the stable region, an increase or decrease in K can cause enough shift in the GH contour to the left or the right to destabilize the system. This can happen because a shaded (unstable) region appears both to the left and the right of the unshaded (stable) region. This phenomenon is called **conditional stability**.

Although absolute stability can often be altered by adjustment of the gain factor alone, other performance criteria such as those concerned with *relative stability* usually require additional compensators.

12.3 GAIN FACTOR COMPENSATION USING *M*-CIRCLES

The gain factor K of G for a *unity feedback* system can be determined for a specific resonant peak M_p by the following procedure which entails drawing the Polar Plot once only.

Step 1: Draw the Polar Plot of $G(\omega)$ for $K = 1$.

Step 2: Calculate Ψ_p, given by

$$\Psi_p = \sin^{-1}\left(\frac{1}{M_p}\right) \tag{12.1}$$

Step 3: Draw a radial line $\overline{AB}$ at an angle Ψ_p below the negative real axis, as shown in Fig. 12-4.

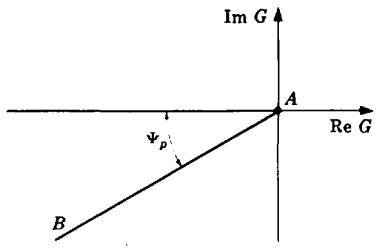


Fig. 12-4

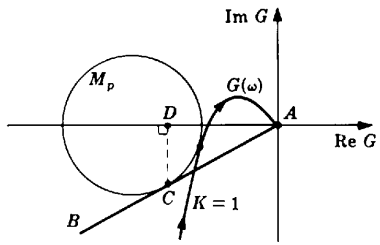


Fig. 12-5

Step 4: Draw the M_p circle tangent to both $G(\omega)$ and line $\overline{AB}$ at C. Then draw a line $\overline{CD}$ perpendicular to the real axis shown in the example Polar Plot shown in Fig. 12-5.

Step 5: Measure the length of line $\overline{AD}$ along the real axis. The required gain factor K for the specified M_p is given by

$$K_{M_p} = \frac{1}{\text{length of line } \overline{AD}} \tag{12.2}$$

If the Polar Plot of G for a gain factor K' other than $K = 1$ is already available, it is not necessary to repeat this plot for $K = 1$. Simply apply Steps 2 through 5 and use the following formula for the gain factor necessary to achieve the specified M_p:

$$K_{M_p} = \frac{K'}{\text{length of line } \overline{AD}} \tag{12.3}$$

12.4 LEAD COMPENSATION

The transfer function for a continuous system lead network, presented in Equation (*6.2*), is

$$P_{\text{lead}} = \frac{s+a}{s+b}$$

where $a < b$. The Polar Plot of P_{Lead} for $0 \le \omega < \infty$ is shown in Fig. 12-6.

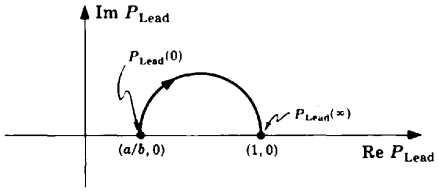


Fig. 12-6

For some systems in which lead compensation is applicable, appropriate choice of the zero at $-a$ and the pole at $-b$ permits an increase in the open-loop gain factor K, providing greater accuracy (and sometimes stability), without adversely affecting transient performance. Conversely, for a given K, transient performance can be improved. In some cases, both steady state and transient response can be favorably modified with lead compensation.

The lead network provides compensation by virtue of its phase lead property in the low-to-medium-frequency range and its negligible attenuation at high frequencies. The low-to-medium-frequency range is defined as the vicinity of the resonant frequency ω_p. Several lead networks may be cascaded if a large phase lead is required.

Lead compensation generally increases the *bandwidth* of a system.

EXAMPLE 12.4. The Polar Plot for

$$GH_1(s) = \frac{K_1}{s(s+p_1)(s+p_2)} \qquad K_1, p_1, p_2 > 0$$

is given in Fig. 12-7. The system is stable and the phase margin ϕ_{PM} is greater than 45°. For a given application, ϕ_{PM} is too large, causing a longer than desired delay time T_d in the system transient response. The steady state error is also too large. That is, the velocity error constant K_v is too small by a factor of $\lambda > 1$. We shall modify this system by a combination of gain factor compensation, to meet the steady state specification, and phase lead compensation, to improve the transient response. Assuming $H(s) = 1$, Equation (*9.12*) yields

$$K_{v1} = \lim_{s \to 0}[sGH_1(s)] = \frac{K_1}{p_1 p_2}$$

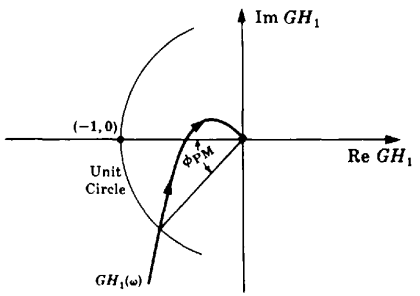


Fig. 12-7

and hence

$$\lambda K_{v1} = \frac{\lambda K_1}{p_1 p_2}$$

Putting $K_2 \equiv \lambda K_1$, the open-loop transfer function becomes

$$GH_2 = \frac{K_2}{s(s+p_1)(s+p_2)}$$

The system represented by GH_2 has the desired velocity constant $K_{v2} = \lambda K_{v1}$.

Let us now consider what would happen to K_{v2} of GH_2 if a lead network were introduced. The lead network acts like an attenuator at low frequencies. That is,

$$\lim_{s \to 0}[sGH_2(s) \cdot P_{\text{Lead}}(s)] = \frac{K_2 a}{p_1 p_2 b} < \lambda K_{v1}$$

since $a/b < 1$. Therefore if a lead network is used to modify the transient response, the gain factor K_1 of GH_1 must be increased $\lambda(b/a)$ times in order to meet the steady state requirement. The gain factor part of the total compensation should therefore be larger than that which would be called for if only the steady state specification has to be met. Hence we modify GH_2, yielding

$$GH_3 = \frac{\lambda K_1 (b/a)}{s(s+p_1)(s+p_2)}$$

As is often the case, increasing the gain factor by an amount as large as $\lambda(b/a)$ times destabilizes the system, as shown in the Polar Plots of GH_1, GH_2, and GH_3 in Fig. 12-8.

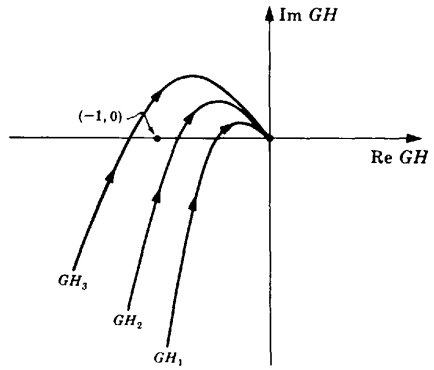


Fig. 12-8

Now let us insert the lead network and determine its effects. GH_3 becomes

$$GH_4 = \frac{\lambda K_1 (b/a)(s+a)}{s(s+p_1)(s+p_2)(s+b)}$$

First, $\lim_{s \to 0}[sGH_4(s)] = \lambda K_{v1}$ convinces us that the steady state specification has been met. In fact, in the very low frequency region we have

$$GH_4(j\omega)\big|_{\omega \text{ very small}} \cong \frac{\lambda K_1}{j\omega(j\omega+p_1)(j\omega+p_2)} = GH_2$$

Hence the GH_4 contour is almost coincident with the GH_2 contour in the very low frequency range.

In the very high frequency region,

$$GH_4(j\omega)\big|_{\omega \text{ very large}} \cong \frac{\lambda K_1 (b/a)}{j\omega(j\omega+p_1)(j\omega+p_2)} = GH_3$$

Therefore the GH_4 contour is almost coincident with GH_3 for very high frequencies.

In the mid-frequency range, where the phase lead property of the lead network substantially alters the phase characteristic of GH_4, the GH_4 contour bends away from the GH_2 and toward the GH_3 locus as ω is increased. This effect is better understood if we write GH_4 in the following form:

$$GH_4(j\omega) = \left[\frac{\lambda K_1(b/a)}{j\omega(j\omega+p_1)(j\omega+p_2)}\right]\cdot\left[\frac{j\omega+a}{j\omega+b}\right]$$

$$= GH_3(j\omega)\cdot P_{\text{Lead}}(j\omega) = GH_3(j\omega)\cdot|P_{\text{Lead}}(j\omega)|\angle\phi(\omega)$$

where $|P_{\text{Lead}}(j\omega)| = \sqrt{(\omega^2+a^2)/(\omega^2+b^2)}$, $\phi(\omega) \equiv \tan^{-1}(\omega/a) - \tan^{-1}(\omega/b)$, $a/b < |P_{\text{Lead}}(j\omega)| < 1$, $0° < \phi(\omega) < 90°$. Therefore the lead network modifies GH_3 as follows. GH_3 is shifted downwards beginning at $GH_3(j\infty)$ in a counterclockwise direction toward GH_2 due to the positive phase contribution of P_{Lead} $[0° < \phi(\omega) < 90°]$. In addition, it is attenuated $[0 < |P_{\text{Lead}}(j\omega)| < 1]$. The resulting Polar Plot for GH_4 is illustrated in Fig. 12-9.

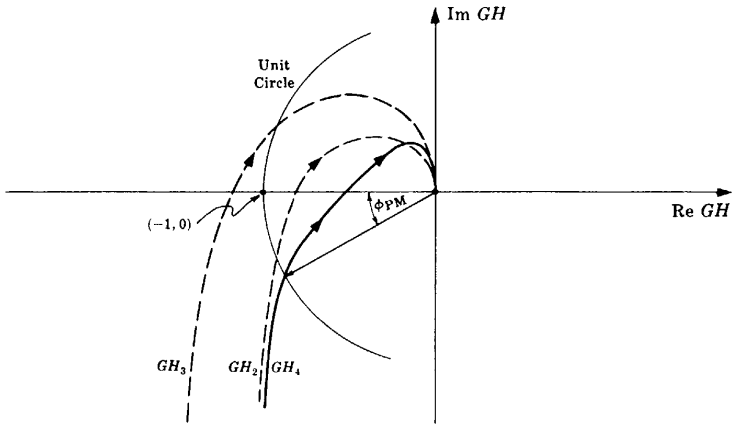


Fig. 12-9

The system represented by GH_4 is clearly stable, and ϕ_{PM} is less than 45°, reducing the delay time T_d of the original system represented by GH_1. By a trial-and-error procedure, the zero at $-a$ and the pole at $-b$ can be chosen such that a specific M_p can be achieved.

A block diagram of the fully compensated system is shown in Fig. 12-10. Unity feedback is shown for convenience only.

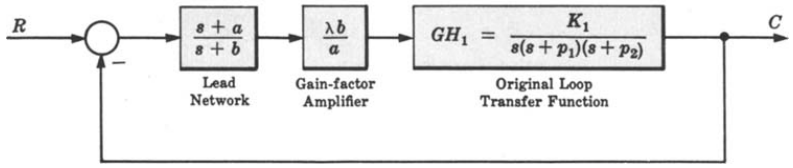


Fig. 12-10

12.5 LAG COMPENSATION

The transfer function for a continuous system lag network, presented in Equation (*6.3*), is

$$P_{\text{Lag}} = \frac{a}{b}\left[\frac{s+b}{s+a}\right]$$

where $a < b$. The Polar Plot of P_{Lag} for $0 \le \omega < \infty$ is shown in Fig. 12-11.

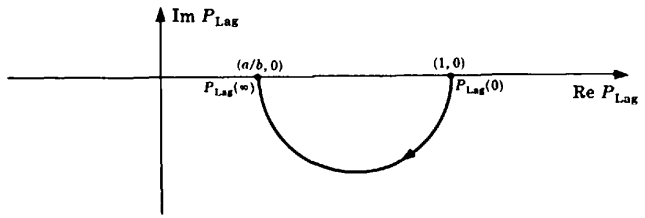


Fig. 12-11

The lag network usually provides compensation by virtue of its attenuation property in the high frequency portion of the Polar Plot, since $P_{\text{Lag}}(0) = 1$ and $P_{\text{Lag}}(\infty) = a/b < 1$. Several lag networks can be cascaded to provide even higher attenuation, if required. The phase lag contribution of the lag network is often restricted by design to the very low frequency range. Several general effects of lag compensation are:

1. The bandwidth of the system is usually decreased.
2. The dominant time constant τ of the system is usually increased, producing a more sluggish system.
3. For a given relative stability, the value of the error constant is increased.
4. For a given value of error constant, relative stability is improved.

The procedure for using lag compensation to improve system performance is essentially the same as that for lead compensation.

EXAMPLE 12.5. Let us redesign the system of Example 12.4 using gain factor plus *lag* compensation. The original open-loop transfer function is

$$GH_1 = \frac{K_1}{s(s+p_1)(s+p_2)}$$

The gain factor compensation transfer function is

$$GH_2 = \frac{\lambda K_1}{s(s+p_1)(s+p_2)}$$

Since $P_{\text{Lag}}(0) = 1$, introduction of the lag network after the steady state criterion has been met by gain factor compensation does not require an additional increase in gain factor.

Introducing the lag network, we get

$$GH_3' = \frac{\lambda K_1(a/b)(s+b)}{s(s+p_1)(s+p_2)(s+a)}$$

Now

$$\lim_{s\to 0}[sGH_3'(s)] = \lambda K_{v1}$$

where $K_{v1} = K_1/p_1p_2$. Therefore the steady state specification is met by GH_3'.

In the very low frequency region,

$$GH_3'(j\omega)|_{\omega \text{ very small}} \cong \frac{\lambda K_1}{j\omega(j\omega+p_1)(j\omega+p_2)} = GH_2(j\omega)$$

Hence GH_3' is almost coincident with GH_2 at very low frequencies, with the lag property of this network manifesting itself in this range.

In the very high frequency region,

$$GH_3'(j\omega)|_{\omega \text{ very large}} \cong \frac{\lambda(a/b)K_1}{j\omega(j\omega+p_1)(j\omega+p_2)} = \lambda(a/b)GH_1(j\omega)$$

Therefore, the GH_3' contour lies above or below the GH_1 contour in the range, if $\lambda > b/a$ or $\lambda < b/a$, respectively. If $\lambda = b/a$, the GH_3' and GH_1 contours coincide.

In the mid-frequency range, the attenuation effect of P_{Lag} increases as ω becomes larger, and there is relatively small phase lag.

The resulting Polar Plot (with $\lambda = b/a$) and a block diagram of the fully compensated system are given in Figs. 12-12 and 12-13.

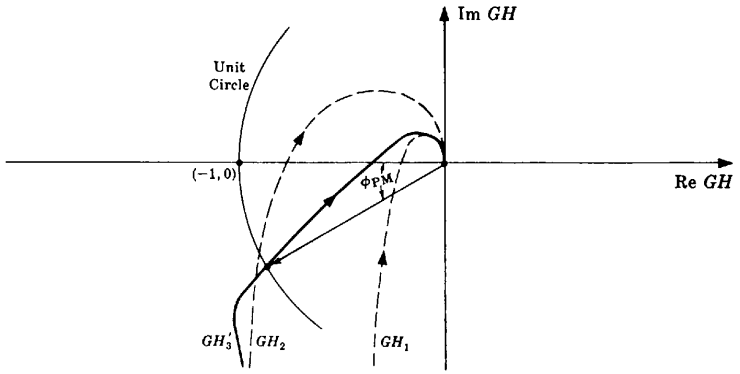


Fig. 12-12

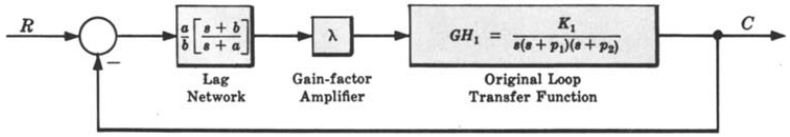


Fig. 12-13

12.6 LAG-LEAD COMPENSATION

The transfer function for a continuous system lag-lead network, presented in Equation (*6.4*), is

$$P_{\text{LL}} = \frac{(s+a_1)(s+b_2)}{(s+b_1)(s+a_2)}$$

where $a_1b_2/b_1a_2 = 1$, $b_1/a_1 = b_2/a_2 > 1$, $a_i, b_i > 0$. The Polar Plot of P_{LL} for $0 \le \omega \le \infty$ is shown in Fig. 12-14.

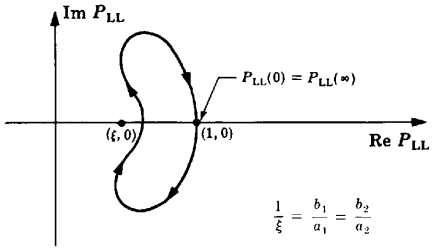


Fig. 12-14

Lag-lead compensation has all of the advantages of both lag compensation and lead compensation, and only a minimum of their usually undesirable characteristics. Satisfaction of many system specifications is possible without the burden of excessive bandwidth and small dominant time constants.

It is not easy to generalize about the application of lag-lead compensation or to prescribe a method for its employment, especially using Nyquist techniques. But, for illustrative purposes, we can describe how it alters the properties of a simple type 2 system in the following example.

EXAMPLE 12.6. The Nyquist Stability Plot for

$$GH = \frac{K}{s^2(s+p_1)} \qquad p_1, K > 0$$

is given in Fig. 12-15. Clearly, the system is unstable, and no amount of gain factor compensation can stabilize it because the contour for $0 < \omega < \infty$ always lies above the negative real axis. Lag compensation is also inapplicable for basically the same reason.

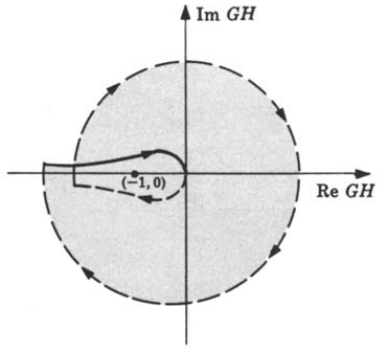


Fig. 12-15

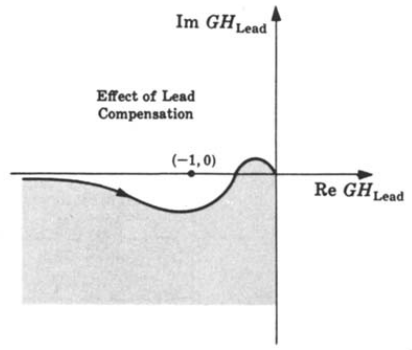


Fig. 12-16

Lead compensation may succeed in stabilizing the system, as shown in Fig. 12-16. But the desired application for the compensated system may call for a lower bandwidth than can be achieved with a lead network.

If a lag-lead network is used, the open-loop transfer function becomes

$$GH_{\text{LL}} = \frac{K(s+a_1)(s+b_2)}{s^2(s+p_1)(s+b_1)(s+a_2)}$$

and the Polar Plot is shown in Fig. 12-17. This system is conditionally stable if the $(-1,0)$ point falls on the real axis in the unshaded region. By trial and error, the parameters of the lag-lead network can be chosen to yield good transient and steady state performance for this previously unstable system, and the bandwidth will be smaller than that of the lead-compensated system. A computer program control system design (CAD) package, or any program that readily generates Polar Plots, can be used to help accomplish this task quickly and effectively.

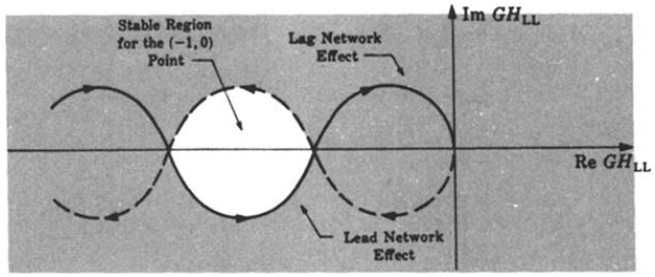


Fig. 12-17

12.7 OTHER COMPENSATION SCHEMES AND COMBINATIONS OF COMPENSATORS

Many other types of physical networks can be used to compensate feedback control systems. Compensation networks can also be implemented in software, as part of the control algorithm in a computer-controlled system. PID controllers are a popular class of such controllers (see Examples 2.14 and 6.7 and Section 10.5).

Combinations of gain factors and lead or lag networks were used as compensators in Examples 12.4 and 12.5, and a lag-lead compensator alone was used in Example 12.6 Other combinations are also feasible and effective, particularly where steady state error requirements cannot be met by gain factor compensation alone. This is often the case when the open-loop transfer function has too few "integrators," that is, denominator terms of the form s^l for continuous systems, or $(z-1)^l$ for discrete-time systems as illustrated in the next example.

EXAMPLE 12.7. Our goal is to determine an appropriate compensator $G_1(z)$ for the digital system shown in Fig. 12-18. The resulting closed-loop system must meet the following performance specifications:

1. Steady state error $e(\infty) = 1 - c(\infty) \le 0.02$, for a unit *ramp* input.
2. Phase margin $\phi_{PM} \ge 30°$.
3. Gain crossover frequency $\omega_1 \ge 10$ rad/sec.*

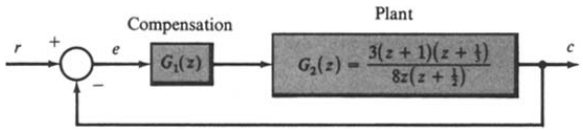


Fig. 12-18

The sampling period for this system is $T = 0.1$ sec (sampling angular frequency $\omega_s = 2\pi/0.1 = 20\pi$ rad/sec).

We note first that the plant is a type 0 system, because there is no "integrator" term of the form $(z-1)^l$ in the denominator of $G_2(z)$ for $l \ge 1$ (see Section 9.8). To meet the first performance specification, it is immediately clear that the overall open-loop system type must be increased by a factor of at least 1, that is, the compensated system must be at least type 1, to achieve a finite steady state error for a unit ramp input. Therefore we add a single pole at $z = 1$, as G_1', as a first step in determining appropriate compensation:

$$G_1'G_2 = \frac{3(z+1)(z+\frac{1}{3})}{8z(z-1)(z+\frac{1}{2})}$$

Now, from the table in Section 9.9, the steady state error for a unit ramp input is $e(\infty) = T/K_v$, and the velocity error constant is $K_v = 3(2)(\frac{4}{3})/8(\frac{3}{2}) = \frac{2}{3}$. Therefore $e(\infty) = 0.15$, which is larger than the value of 0.02 required by performance specification 1.

The next obvious question is whether the addition of gain factor compensation would be sufficient to complete the design. This would require a gain increase by at least a factor of $\lambda = 0.1/(0.02)(\frac{2}{3}) = \frac{15}{2}$, yielding

$$G_1''G_2 = \frac{15}{2}G_1G_2' = \frac{45(z+1)(z+\frac{1}{3})}{16z(z-1)(z+\frac{1}{2})}$$

To check the remaining performance criteria (2 and 3) the gain crossover frequency ω_1 and phase margin ϕ_{PM} can be evaluated from their defining equations in Section 11.11. We have

$$\phi_{PM} = [180 + \arg G_1''G_2(\omega_1)] \text{ degrees}$$

and ω_1 satisfies the equation

$$|G_1''G_2(\omega_1)| = 1$$

Now, ω_1 and ϕ_{PM} could be determined graphically from a Nyquist Stability Plot of $G_1''G_2$, as illustrated in Fig. 11-16. But a less difficult task is to solve for ω_1 and ϕ_{PM} from their defining equations, preferably using a computer

*See Problem 12.16 for further discussion of this performance specification and its relationship to system *bandwidth* BW.

program capable of complex numerical calculations. This can be done by first substituting $e^{j\omega T}$ for z in $G_1''G_2(z)$, using the Polar Form, Euler Form, and/or Complex Form substitutions [Equations (*11.2*) through (*11.4*)], and then solving for $\omega_1 T$ such that $|G_1''G_2| = 1$. Trial-and-error solution for $\omega_1 T$ can be helpful in this regard, which we used to find $\omega_1 T = 2.54$ rad after several trials, resulting in $G_1''G_2(\omega_1) = -0.72 + j0.7$ and

$$\phi_{PM} = \left[180° - \tan^{-1}\left(\frac{0.7}{-0.72}\right)\right] = -44.4°$$

Clearly, $\omega_1 = 2.54/0.1 = 25.4 > 10$ rad/sec satisfies performance specification 3, but *not* the phase margin requirement 2, because $\phi_{PM} = -44.4 \not\ge 30°$, the negative phase margin also indicating that the closed-loop system with $G_1''G_2$ is unstable.

Introduction of a *lag* compensator might solve the remaining constraint, because it increases the phase margin without affecting the steady state error. The transfer function of a digital lag compensator was given in Example 6.12, Equation (*6.11*), as

$$P_{Lag}(z) = \left(\frac{1-p_c}{1-z_c}\right)\left[\frac{z-z_c}{z-p_c}\right] \qquad (12.4)$$

where $z_c < p_c$. Note that $P_{Lag}(1) = P_{Lag}(e^{j0}) = 1$, which explains why the lag network does not affect the steady state response of this type 1 system. The Polar Plot of P_{Lag} is shown in Fig. 12-26.

The problem now is to choose appropriate values of z_c and p_c to render $\phi_{PM} \ge 30°$ and $\omega_1 \ge 10$ rad/sec. Again, we accomplished this readily by trial and error, using a computer to evaluate the simultaneous solution for z_c and p_c of the two relations $|G_1''' G_2(10)| = 1$ and

$$\phi_{PM} = [180 + \arg G_1''' G_2(10)] \ge 30°$$

where $G_1''' G_2 = P_{Lag}(G_1''G_2)$. These equations have multiple solutions and, often, good choices for p_c and z_c are values close to 1, because P_{Lag} then has minimal effect on the phase of $G_1''G_2$ at higher frequencies. The pole and zero of P_{Lag} effectively cancel each other at high frequencies when their values are close to 1. After several trials, we obtained $a = 0.86$ and $b = 0.97$, and a final compensator:

$$G_1(z) \equiv G_1'''(z) = \frac{1.59(z-0.86)}{(z-1)(z-0.97)}$$

The resulting Polar Plot (for $0 < \omega < \pi$) for the compensated system G_1G_2 is shown with $\phi_{PM} > 30°$ in Fig. 12-19.

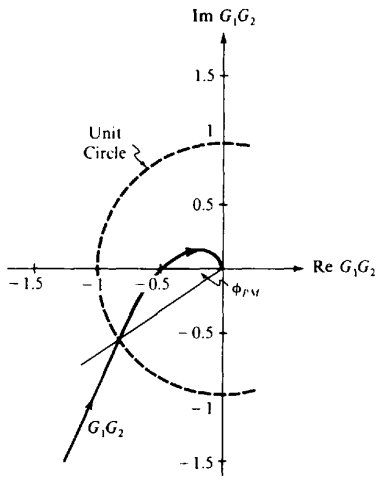


Fig. 12-19

Example 12.7 is reworked by root-locus techniques in Example 14.5, and also by Bode methods in Example 16.6, the latter solution using the *w*-transform introduced in Section 10.7.

Solved Problems

GAIN FACTOR COMPENSATION

12.1. Consider the open-loop transfer function $GH = -3/(s+1)(s+2)$. Is the system represented by GH stable or unstable?

Unstable. The characteristic equation is determined from $1 + GH = 0$ and is given by $s^2 + 3s - 1 = 0$. Since all the coefficients do not have the same sign, the system is unstable (see Problem 5.27).

12.2. Determine the minimum value of gain factor to stabilize the system of the previous problem.

Let GH be written as $GH = K/(s+1)(s+2)$. Then the characteristic equation is $s^2 + 3s + 2 + K = 0$ and the Routh table (see Section 5.3) is

s^2	1	$(2+K)$
s^1	3	0
s^0	$(2+K)$	

Hence the minimum gain factor for stability is $K = -2 + \epsilon$, where ϵ is any small positive number.

12.3. The solution of the previous problem also tells us that the system of Problems 12.1 and 12.2 is stable for all $K > -2$. Sketch Polar Plots of this system, superimposed on the same coordinate axes, for $K_1 = -3$ and $K_2 = -1$. What general comments can you make about the transient response of the stable system? Assume it is a unity feedback system.

The required Polar Plots are shown in Fig. 12-20. The M-circle tangent to the plot for $K = -1$ has infinite radius; thus $M_p = 1$. This means that the peak overshoot is zero (no overshoot), and the system is either critically damped or overdamped.

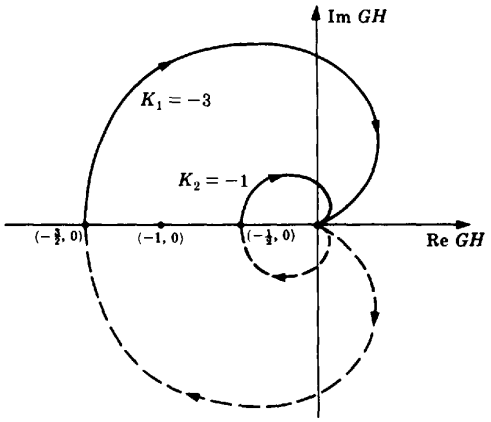


Fig. 12-20

12.4. Is the system represented by the characteristic equation $s^3 + 3s^2 + 3s + 1 + K = 0$ ever conditionally stable? Why?

Yes. The gain factor range for stability of this system was determined in Example 5.3 as $-1 < K < 8$. Since both limits are finite, an increase in the gain factor above 8 or a decrease below -1 destabilizes the system.

12.5. Determine the gain factor K of a unity feedback system whose open-loop transfer function is given by $G = K/(s+1)(s+2)$ for a resonant peak specified by $M_p = 2$.

From Equation (*12.1*) we have $\Psi_p = \sin^{-1}(\frac{1}{2}) = 30°$. The line $\overline{AB}$ drawn at an angle of 30° below the negative real axis is shown in Fig. 12-21, a replica of Fig. 12-20 for $K = -1$.

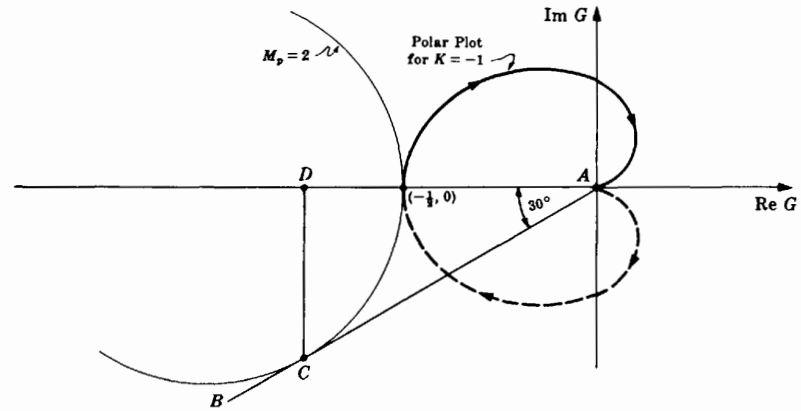


Fig. 12-21

The circle denoted by $M_p = 2$ has been drawn tangent to both $\overline{AB}$ and the Polar Plot of $K = -1$. Using the scale of this Polar Plot, line $\overline{AD}$ has a length equal to 0.76. Therefore Equation (*12.3*) yields

$$K_{M_p} = \frac{K'}{\text{length of } \overline{AD}} = \frac{-1}{0.76} = -1.32$$

It is also possible to compute a *positive* value of gain for $M_p = 2$ from a Polar Plot of $G(s)$ for any positive value of K. The Polar Plot for $K = 1$ is the same as that in Fig. 12-21, but rotated by 180°.

MISCELLANEOUS COMPENSATION

12.6. What kind of compensation is possible for a system whose Polar Plot is given by Fig. 12-22?

Lead, lag-lead, and simple gain factor compensation are capable of stabilizing the system and improving the relative stability.

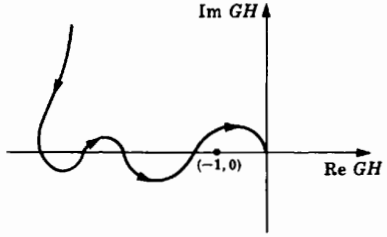


Fig. 12-22

12.7. Consider the unity feedback system whose open-loop transfer function is given by

$$G = \frac{K_1}{s(s+a)} \qquad a, K_1 > 0$$

How would the inclusion of a minor feedback loop with a transfer function $K_2 s$ ($K_2 > 0$), as shown in the block diagram in Fig. 12-23, affect the transient and steady state performance of the system?

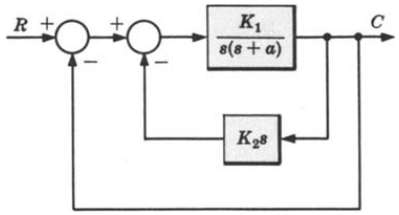


Fig. 12-23

Combining the blocks in the inner loop yields a new unity feedback system with an open-loop transfer function

$$G' = \frac{K_1}{s(s + a + K_1 K_2)}$$

The Polar Plots for G and G' are sketched in Fig. 12-24.

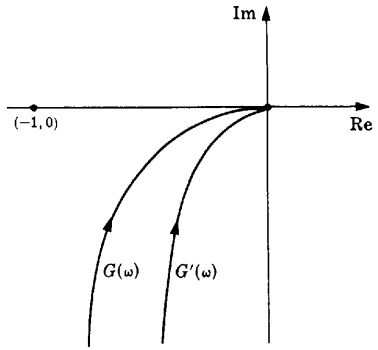


Fig. 12-24

The phase margin is clearly larger for the two-loop feedback system G'. Hence the peak overshoot is smaller, or the damping ratio is larger, and the transient response is superior to that of the uncompensated system. The steady state performance however, is generally slightly worse. For a unit step input the steady state error is zero, as for any type 1 system. But the steady state error for a unit ramp or velocity input is larger [see Equations (*9.4*) and (*9.5*)]. The compensation scheme illustrated by this problem is called *derivative* or *tachometric feedback*, and the control algorithm is *derivative* (D) control.

12.8. Determine a type of compensator that yields a phase margin of approximately 45° when added to the fixed system components defined by

$$GH = \frac{4}{s(s^2 + 3.2s + 64)}$$

An additional requirement is that the high-frequency response of the compensated system is to be approximately the same as that of the uncompensated system.

The Polar Plot for GH is sketched in Fig. 12-25. It is very close to the negative imaginary axis for almost all values of ω.

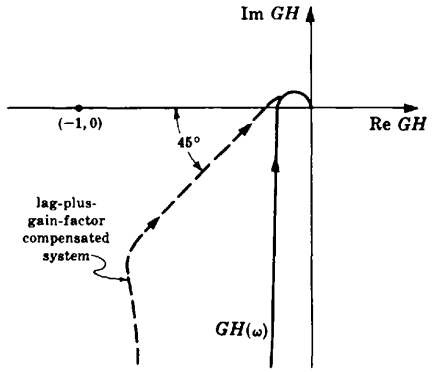


Fig. 12-25

The phase margin is almost 90°, and either an increase in gain factor and/or a lag compensator is capable of satisfying the phase margin requirement. But since the lag network may be designed to provide attenuation at high frequencies and lag in the low-frequency range, a combination of both would be ideal and sufficient (see Example 12.5), as shown in Fig. 12-25. Of course a lag-plus-gain factor compensator is not *necessary* for meeting the design requirements. There are probably an infinite number of different networks or transfer functions capable of satisfying these specifications. The lag network and amplifier, however, are *convenient* due to their standardization, availability, and ease of synthesis.

12.9. Outline the design of a servomechanism capable of following a constant velocity input with zero steady state error and approximately 25% maximum overshoot in the transient state. The fixed plant is given by $G_2 = 50/s^2(s + 5)$.

Since the plant is type 2, it is capable of following a constant velocity input with zero steady state error (see Chapter 9). However, the closed-loop system is unstable for any value of gain factor (see Example 12.6). Since no demands on bandwidth have been made, lead compensation should be sufficient (again see Example 12.6) to stabilize the system and meet the transient specification. But two lead networks in series are probably required because the phase margin of the unstable system is negative, and 25% overshoot is equivalent to about +45° phase margin. Most standard lead networks have a maximum phase lead of approximately 54° (see Fig. 16-2).

Detailed design would be very tedious using Nyquist analysis, if performed manually, because the Polar Plot usually must be drawn in some detail several times before converging to a satisfactory solution. If a computer is not available to facilitate this process, this problem may be solved much more easily using the design methods introduced in Chapters 14, 16, and 18. Actually, two compensating lead networks, each with a transfer function of approximately $P_{\text{Lead}} = (s + 3)/(s + 20)$, would satisfy the specifications. If the maximum steady state acceleration error were also specified, a preamplifier would be required with the lead networks. For example, if $K_a = 50$, then a preamplifier of gain $5(20/3)^2$ would be needed. This preamplifier should be placed *between* the two lead networks to prevent, or minimize, loading effects (see Section 8.7).

12.10. Outline a design for a unity feedback system with a plant given by

$$G_2 = \frac{2000}{s(s + 5)(s + 10)}$$

and the performance specifications:

(1) $\phi_{\text{PM}} \cong 45°$.

(2) $K_v = 50$.

(3) The bandwidth BW of the compensated system must be approximately equal to or not much greater than that of the uncompensated system, because high-frequency "noise" disturbances are present under normal operating conditions.

(4) The compensated system should not respond sluggishly; that is the predominant time constant τ of the system must be maintained at a value approximately the same as that of the uncompensated system.

A simple calculation clearly shows that the uncompensated system is unstable (e.g., try the Routh test). Therefore compensation is mandatory. But due to the stringent nature of the specifications, a detailed design for this system using Nyquist techniques requires too much effort, if done manually. The techniques of the next few chapters provide a much simpler solution. However, analysis of the problem statement indicates the kind of compensation needed.

For G_2, $K_v = \lim_{s\to 0} sG_2(s) = 40$. Therefore satisfaction of (2) requires a gain compensation of 5/4. But an increase in gain only makes the system more unstable. Therefore additional compensation is necessary. Lead compensation is probably inadequate due to (3), and lag compensation is not possible due to (4). Thus it appears that a lag-lead network and an amplifier would most likely satisfy all criteria. The lag portion of the lag-lead network would satisfy (3), and the lead portion (4) and (1).

12.11. What is the effect on the Polar Plot of the system

$$GH = \frac{\prod_{i=1}^{m}(s+z_i)}{\prod_{i=1}^{n}(s+p_i)}$$

where $m \le n$, $0 < z_i < \infty$, $0 \le p_i < \infty$, when k finite nonzero poles are included in GH, in addition to the original n poles?

For low frequencies the Polar Plot is modified in magnitude only, since

$$\lim_{s\to 0} GH' = \lim_{s\to 0}\left[\frac{\prod_{i=1}^{m}(s+z_i)}{\prod_{i=1}^{n+k}(s+p_i)}\right] = \frac{\prod_{i=1}^{m} z_i}{\prod_{i=1}^{n+k} p_i} = \left(\frac{1}{\prod_{i=1}^{k} p_i}\right)\lim_{s\to 0} GH$$

For high frequencies addition of k poles reduces the phase angle of GH by $k\pi/2$ radians, since

$$\lim_{\omega\to\infty} \arg GH'(\omega) = \lim_{\omega\to\infty}\left[\sum_{i=1}^{m}\tan^{-1}\left(\frac{\omega}{z_i}\right) - \sum_{i=1}^{n+k}\tan^{-1}\left(\frac{\omega}{p_i}\right)\right]$$

$$= \frac{m\pi}{2} - \frac{(n+k)\pi}{2} = \lim_{\omega\to\infty}\arg GH - \frac{k\pi}{2}$$

Therefore the portion of the Polar Plot near the origin is rotated clockwise by $k\pi/2$ degrees when k poles are added.

12.12. Draw the Polar Plot of the digital lag compensator given by Equation (*12.4*):

$$P_{\text{Lag}}(z) = \left(\frac{1-p_c}{1-z_c}\right)\left[\frac{z-z_c}{z-p_c}\right] \qquad z_c < p_c$$

Let $z_c = 0.86$ and $p_c = 0.97$, to simplify the task.

At $\omega = 0$, $P_{\text{Lag}} = P_{\text{Lag}}(e^{j0T}) = P_{\text{Lag}}(1) = 1$. At $\omega T = \pi$,

$$P_{\text{Lag}}(e^{j\pi}) = \left(\frac{1-p_c}{1-z_c}\right)\left[\frac{-1-z_c}{-1-p_c}\right] = \frac{1-z_cp_c-(p_c-z_c)}{1-z_cp_c+(p_c-z_c)} \equiv c = 0.2$$

At a few intermediate values, $P_{\text{Lag}}(e^{j\pi/4}) \approx 0.02 - j0.03$ and $P_{\text{Lag}}(e^{j\pi/2}) \approx 0.2 - j0.012$. The resulting Polar Plot, for $0 \le \omega T \le \pi$ radians is shown in Fig. 12-26. It is instructive to compare this Polar Plot of the digital lag compensator with its continuous-time equivalent in Fig. 12-11.

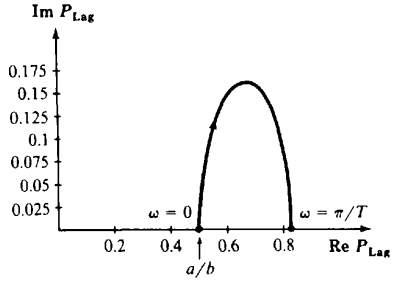


Fig. 12-26

12.13. Draw the Polar Plot of the particular *digital lead* compensator:

$$P_{\text{Lead}}(z) = \left(\frac{a}{b}\right)\left[\frac{1-e^{-bT}}{1-e^{-aT}}\right]\left[\frac{z-e^{-aT}}{z-e^{-bT}}\right]$$

where $a < b$.

We have

$$P_{\text{Lead}}(e^{j0T}) = P_{\text{Lead}}(1) = \left(\frac{a}{b}\right)\left(\frac{1-e^{-bT}}{1-e^{-aT}}\right)\left(\frac{1-e^{-aT}}{1-e^{-bT}}\right) = \frac{a}{b} < 1$$

The remainder of the plot has been drawn by computer, by evaluating $P_{\text{Lead}}(1\angle\phi)$ for values of the angle ϕ in the range $0 < \phi \le \pi$ radians, for specific values $a = 1$ and $b = 2$. The result is given in Fig. 12-27, which should be compared with Fig. 12-6, the Polar Plot of a continuous system lead network.

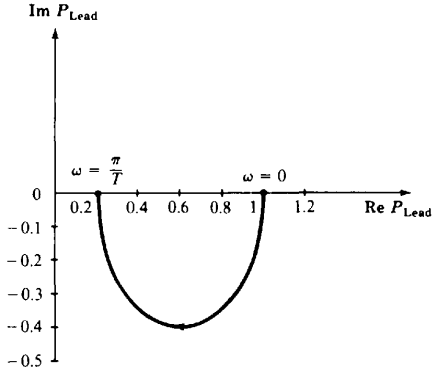


Fig. 12-27

This form of the general digital lead compensator, given in Equation (*6.9*), has a gain factor

$$K_{\text{Lead}} = \frac{a}{b}\left[\frac{1-e^{-bT}}{1-e^{-aT}}\right]$$

This compensator is a direct digital analog of the continuous lead compensator $P_{\text{Lead}} = (s+a)/(s+b)$, in which the zeros and poles at $-a$ and $-b$ in the s-plane have been transformed directly into zeros and poles in z-plane $z_c = e^{-aT}$ and $p_c = e^{-bT}$, and the steady state gain (at $\omega = 0$) has been preserved as a/b.

12.14. The closed-loop continuous system with both gain factor and lead compensation shown in Fig. 12-28 is stable, with a damping ratio $\zeta \approx 0.7$ and dominant time constant $\tau \approx 4.5$ sec (see

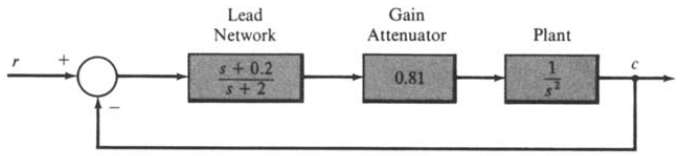


Fig. 12-28

Sections 4.13 and 10.4). Redesign this system, replacing the controller (including summing junction) with a digital computer, and any other needed components for analog-digital data conversion. The new system should have approximately the same dynamic characteristics.

The sampling rate of the digital components must be sufficiently fast to reproduce the signals accurately. The natural frequency ω_n is estimated from Equation (*10.7*) as $\omega_n \approx 1/\xi\tau = 1/(0.7)(4.5) = 0.317$ rad/sec. For a continuous system with this ω_n, a safe angular sampling frequency $\omega_s \approx 20\omega_n = 6.35 \approx 2\pi$ rad/sec, equivalent to $f_s = 1$ Hz, because $\omega_s = 2\pi f_s$. Therefore we choose $T = 1$ sec.

We now replace the continuous lead compensator by the digital lead compensator given in Problem 12.13:

$$P_{\text{Lead}}(z) = \left(\frac{a}{b}\right)\left(\frac{1-e^{-bT}}{1-e^{-aT}}\right)\left[\frac{z-e^{-aT}}{z-e^{-bT}}\right]$$

$$\simeq 0.55\left[\frac{z-0.82}{z-0.14}\right]$$

where $a = 0.2$ and $b = 2$ from Fig. 12-28. The factor of 0.55 can be obtained with the gain factor compensator for the continuous system, $K = 0.81$, yielding an overall factor of 0.55 (0.81) = 0.45. The resulting design also needs samplers in the feedback and the input paths, and a zero-order hold in the forward path, all as shown in Fig. 12-29.

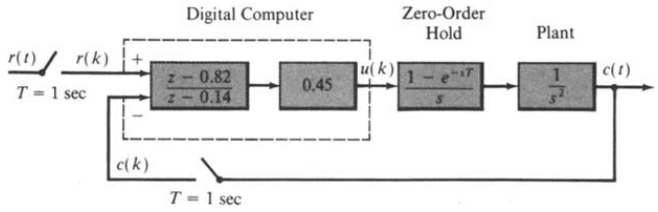


Fig. 12-29

The digital transfer function $P_{\text{Lead}}(z)$ can be implemented for digital computation as a difference equation between the input and output of P_{Lead}, using the methods described in Section 4.9. That is, write $P_{\text{Lead}}(z)$ as a function of z^{-1} instead of z, and treat z^{-1} as a unit time-shift operator. Combining the gain factor 0.45 with P_{Lead}, we obtain

$$0.48P_{\text{Lead}} = \frac{0.45 - 0.39z^{-1}}{1 - 0.14z^{-1}} \equiv \frac{u(k)}{r(k) - c(k)}$$

Then, cross-multiplying terms and letting $z^{-1}u(k) = u(k-1)$, etc., we obtain the desired difference equation:

$$u(k) = 0.14u(k-1) + 0.45[r(k) - c(k)] - 0.39[r(k-1) - c(k-1)]$$

12.15. Digitize the remaining continuous components in Fig. 12-29 and compare the Polar Plot of: (*a*) the original continuous plant without compensation, $G_2(s) = 1/s^2$, (*b*) the compensated system of Fig. 12-28, $G_1G_2(s)$, and (*c*) the digital system of Fig. 12-30, $G_1G_2(z)$.

The combination of the zero-order hold and the plant $G_2(s) = 1/s^2$ can be digitized using Equation (*6.9*):

$$G_2'(z) = \left(\frac{z-1}{z}\right)\mathcal{Z}\left\{\mathcal{L}^{-1}\left(\frac{1}{s^2}\right)\Big|_{t=kT}\right\}$$

$$= \frac{T^2}{2}\left(\frac{z+1}{(z-1)^2}\right) = \frac{0.5(z+1)}{(z-1)^2}$$

The closed-loop discrete-time equivalent system is shown in Fig. 12-30.

Nyquist Stability Plots (not shown) would indicate that the compensated systems are absolutely stable. To check relative stability, the Polar Plots of the three systems are shown superimposed in Fig. 12-31, for $\omega > 0$ only. The phase margin of $G_1G_2(s)$ is $\phi_{\text{PM}} \simeq 53°$, a substantial improvement over that of $G_2(s)$. The Polar Plots for $G_1G_2(s)$ and $G_1G_2(z)$ are quite similar, over a wide range of ω, and the phase margin for $G_1G_2(z)$ is still quite good, $\phi_{\text{PM}} \simeq 37°$.

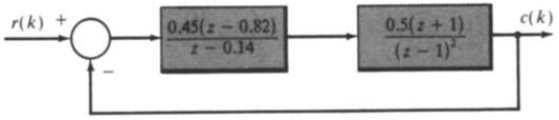


Fig. 12-30

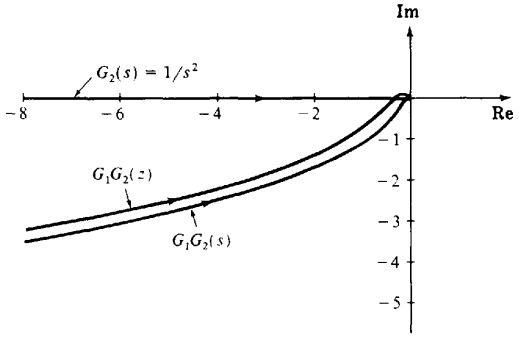


Fig. 12-31

12.16. Determine the closed-loop system bandwidth BW of the compensated system designed in Example 12.7.

Mathcad

Performance specification 3 was given in terms of the gain crossover frequency ω_1, as $\omega_1 \geq 10$ rad/sec. This may appear somewhat unrealistic, or artificial, given that a specific phase margin $\phi_{\text{PM}} = [180 + \arg GH(\omega_1)]$ degrees was also given in performance specification 2. Actually, the *bandwidth* (BW) of the closed-loop system would be the more likely frequency of interest in control system design. (These design criteria are discussed in Chapter 10.) However, as noted in Section 10.4, it is often the case that ω_1 is a good approximation of the closed-loop system bandwidth BW, when it is given its common interpretation as the range of frequencies over which the magnitude ratio of the system, which in this case means $|C/R|$, does not fall more than 3 db from its steady state value, at $\omega = 0$ ($z = 1$). For this problem

$$G_1 = \frac{1.59(z-0.86)}{(z-1)(z-0.97)}$$

$$G_2 = \frac{3(z+1)\left(z+\frac{1}{3}\right)}{8z\left(z+\frac{1}{2}\right)}$$

$$\frac{C}{R} = \frac{G_1G_2}{1+G_1G_2}$$

We easily find that

$$\lim_{\omega \to 0}\left(\frac{C}{R}\right) = \lim_{z \to 1}\left(\frac{C}{R}\right) = 1$$

Now, 3 db down from 1 is 0.707 [see Equation (*10.5*)]. Therefore the BW is the frequency ω_{BW} that satisfies the equation:

$$\left|\frac{C}{R}(\omega_{BW})\right| = 0.707$$

We quickly obtain the solution $\omega_{BW} = 10.724$ rad/sec by trial and error using a computer to evaluate the magnitude ratio at a few values of ω in the vicinity of $\omega_1 = 10$. Thus the approximation $\omega_1 \simeq \omega_{BW}$ is confirmed as a good one for the problem solved in Example 12.7.

Supplementary Problems

12.17. Determine a positive value of gain factor K when $M_p = 2$ for the system of Problem 12.5.

12.18. Prove Equation (*12.1*).

12.19. Prove Equations (*12.2*) and (*12.3*).

12.20. Design a compensator which yields a phase margin of approximately 45° for the system defined by $GH = 84/s(s + 2)(s + 6)$.

12.21. Design a compensator which yields a phase margin of about 40° and a velocity constant $K_v = 40$ for the system defined by $GH = (4 \times 10^5)/s(s + 20)(s + 100)$.

12.22. What kind of compensation can be used to yield a maximum overshoot of 20% for the system defined by $GH = (4 \times 10^4)/s^2(s + 100)$?

12.23. Show that the addition of k finite zeros ($z_i \neq 0$) to the system of Problem 12.11 rotates the high-frequency portion of the Polar Plot by $k\pi/2$ radians in the counterclockwise direction.

Answers to Some Supplementary Problems

12.17. $K = 31.2$

12.18. $P_{\text{Lead}} = \dfrac{s + 30}{s + 120}$

12.21. $P_{\text{Lead}} = \dfrac{s + 20}{s + 100}$, no preamplifier required

12.22. Lag-lead, and possibly lead plus gain factor compensation.

Chapter 13

Root-Locus Analysis

13.1 INTRODUCTION

It was shown in Chapters 4 and 6 that the poles of a transfer function can be displayed graphically in the s-plane or z-plane by means of a pole-zero map. An analytical method is presented in this chapter for displaying the location of the poles of the closed-loop transfer function

$$\frac{G}{1 + GH}$$

as a function of the gain factor K (see Sections 6.2 and 6.6) of the open-loop transfer function GH. This method, called *root-locus analysis*, requires that only the location of the poles and zeros of GH be known, and does not require factorization of the characteristic polynomial.

Root-locus techniques permit accurate computation of the time-domain response in addition to yielding readily available frequency response information.

The following discussion of root-locus analysis applies identically to continuous systems in the s-plane and discrete-time systems in the z-plane.

13.2 VARIATION OF CLOSED-LOOP SYSTEM POLES: THE ROOT-LOCUS

Consider the canonical feedback control system given in Fig. 13-1. The closed-loop transfer function is

$$\frac{C}{R} = \frac{G}{1 + GH}$$

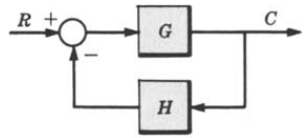


Fig. 13-1

Let the open-loop transfer function GH be represented by

$$GH \equiv \frac{KN}{D}$$

where N and D are finite polynomials in the complex variable s or z and K is the open-loop gain factor. The closed-loop transfer function then becomes

$$\frac{C}{R} = \frac{G}{1 + KN/D} = \frac{GD}{D + KN}$$

The closed-loop poles are roots of the characteristic equation

$$D + KN = 0 \qquad (13.1)$$

In general the location of these roots in the s-plane or z-plane changes as the open-loop gain factor K is varied. A locus of these roots plotted in the s-plane or z-plane as a function of K is called a **root-locus**.

For K equal to zero, the roots of Equation (*13.1*) are the roots of the polynomial D, which are the same as the poles of the open-loop transfer function GH. If K becomes very large, the roots approach

those of the polynomial N, the open-loop zeros. Thus, as K is increased from zero to infinity, the loci of the closed-loop poles originate from the open-loop poles and terminate at the open-loop zeros.

EXAMPLE 13.1. Consider the continuous system open-loop transfer function

$$GH = \frac{KN(s)}{D(s)} = \frac{K(s+1)}{s^2+2s} = \frac{K(s+1)}{s(s+2)}$$

For $H = 1$, the closed-loop transfer function is

$$\frac{C}{R} = \frac{K(s+1)}{s^2+2s+K(s+1)}$$

The closed-loop poles of this system are easily determined by factoring the denominator polynomial:

$$p_1 = -\tfrac{1}{2}(2+K) + \sqrt{1+\tfrac{1}{4}K^2}$$

$$p_2 = -\tfrac{1}{2}(2+K) - \sqrt{1+\tfrac{1}{4}K^2}$$

The locus of these roots plotted as a function of K (for $K > 0$) is shown in the s-plane in Fig. 13-2. As observed in the figure, this root-locus has two *branches*: one for a closed-loop pole which moves from the open-loop pole at the origin to the open-loop zero at -1, and from the open-loop pole at -2 to the open-loop zero at $-\infty$.

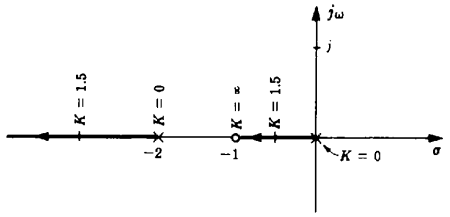


Fig. 13-2

In the example above, the root-locus is constructed by factoring the denominator polynomial of the system closed-loop transfer function. In the following sections, techniques are described which permit construction of root-loci without the need for factorization.

13.3 ANGLE AND MAGNITUDE CRITERIA

In order for a branch of a root-locus to pass through a particular point p_1 in the complex plane, it is necessary that p_1 be a root of the characteristic Equation (*13.1*) for some real value of K. That is,

$$D(p_1) + KN(p_1) = 0 \qquad (13.2)$$

or, equivalently,

$$GH = \frac{KN(p_1)}{D(p_1)} = -1 \qquad (13.3)$$

Therefore the complex number $GH(p_1)$ must have a phase angle of $180° + 360l°$, where l is an arbitrary integer. Thus we have the **angle criterion**

$$\arg GH(p_1) = 180° + 360l° = (2l+1)\pi \text{ radians} \qquad l = 0, \pm 1, \pm 2, \ldots \qquad (13.4a)$$

which can also be written as

$$\arg\left[\frac{N(p_1)}{D(p_1)}\right] = \begin{cases} (2l+1)\pi \text{ radians} & \text{for } K > 0 \\ 2l\pi \text{ radians} & \text{for } K < 0 \end{cases} \qquad l = 0, \pm 1, \pm 2, \ldots \qquad (13.4b)$$

In order for p_1 to be a closed-loop pole of the system, on the root-locus, it is necessary that Equation (*13.3*) be satisfied with regard to *magnitude* in addition to phase angle. That is, K must have the particular value that satisfies the **magnitude criterion**: $|GH(p_1)| = 1$, or

$$|K| = \left|\frac{D(p_1)}{N(p_1)}\right| \qquad (13.5)$$

The angle and magnitude of GH at any point in the complex s- or z-plane can be determined graphically as described in Sections 4.12 and 6.5. In this way, it is possible to construct the root-locus manually by a trial-and-error procedure of testing points in the complex plane. That is, the root-locus is drawn through all points which satisfy the angle criterion, Equation (*13.4b*), and the magnitude criterion is used to determine the values of K at points along the loci. Digital computer programs for routinely plotting root-loci are widely available. However, manual construction is simplified considerably, using certain shortcuts or construction rules as described in the following sections.

13.4 NUMBER OF LOCI

The number of loci, that is, the number of branches of the root-locus, is equal to the number of poles of the open-loop transfer function GH (for $n \geq m$).

EXAMPLE 13.2. The open-loop transfer function of the discrete-time system $GH(z) = K(z+\frac{1}{2})/z^2(z+\frac{1}{4})$ has three poles. Hence there are three loci in the root-locus plot.

13.5 REAL AXIS LOCI

Those sections of the root-locus on the real axis in the complex plane are determined by counting the total number of finite poles and zeros of GH to the right of the points in question. The following rule depends on whether the open-loop gain factor K is positive or negative.

Rule for $K > 0$

Points of the root-locus on the real axis lie to the left of an *odd* number of finite poles and zeros.

Rule for $K < 0$

Points of the root-locus on the real axis lie to the left of an *even* number of finite poles and zeros.

If no points on the real axis lie to the left of an odd number of finite poles and zeros, then no portion of the root-locus for $K > 0$ lies on the real axis. A similar statement is true for $K < 0$.

EXAMPLE 13.3. Consider the pole-zero map of an open-loop transfer function GH shown in Fig. 13-3. Since all the points on the real axis between 0 and -1 and between -1 and -2 lie to the left of an odd number of finite poles and zeros, these points are on the root-locus for $K > 0$. The portion of the real axis between $-\infty$ and -4 lies to the left of an odd number of finite poles and zeros; hence these points are also on the root-locus for $K > 0$. All portions of the root-locus for $K > 0$ on the real axis are illustrated in Fig. 13-4. All remaining portions of the real axis, that is, between -2 and -4 and between 0 and ∞, lie on the root-locus for $K < 0$.

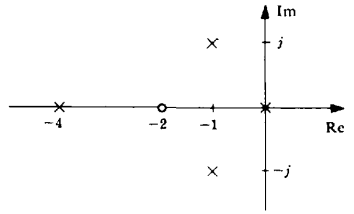


Fig. 13-3

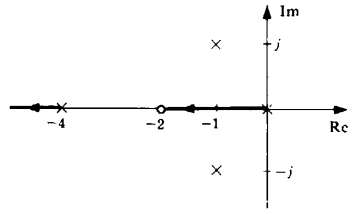


Fig. 13-4

13.6 ASYMPTOTES

For large distances from the origin in the complex plane, the branches of a root-locus approach a set of straight-line asymptotes. These asymptotes emanate from a point in the complex plane on the real axis called the **center of asymptotes** σ_c given by

$$\sigma_c = -\frac{\sum_{i=1}^{n} p_i - \sum_{i=1}^{m} z_i}{n-m} \tag{13.6}$$

where $-p_i$ are the poles, $-z_i$ are the zeros, n is the number of poles, and m the number of zeros of GH.

The angles between the asymptotes and the real axis are given by

$$\beta = \begin{cases} \dfrac{(2l+1)180}{n-m} \text{ degrees} & \text{for } K>0 \\ \dfrac{(2l)180}{n-m} \text{ degrees} & \text{for } K<0 \end{cases} \tag{13.7}$$

for $l = 0, 1, 2, \ldots, n-m-1$. This results in a number of asymptotes equal to $n-m$.

EXAMPLE 13.4. The center of asymptotes for $GH = K(s+2)/s^2(s+4)$ is located at

$$\sigma_c = -\frac{4-2}{2} = -1$$

Since $n-m = 3-1 = 2$, there are two asymptotes. Their angles with the real axis are 90° and 270°, for $K>0$, as shown in Fig. 13-5.

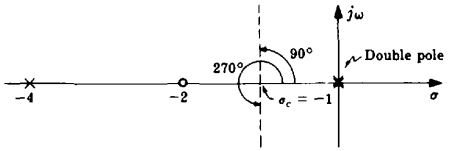


Fig. 13-5

13.7 BREAKAWAY POINTS

A **breakaway point** σ_b is a point on the real axis where two or more branches of the root-locus depart from or arrive at the real axis. Two branches leaving the real axis are illustrated in the root-locus plot in Fig. 13-6. Two branches coming onto the real axis are illustrated in Fig. 13-7.

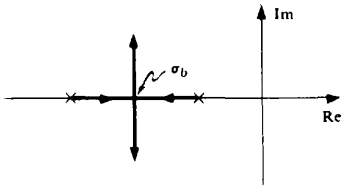


Fig. 13-6

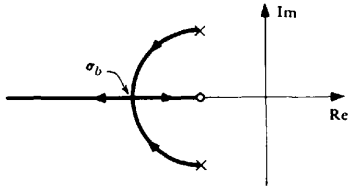


Fig. 13-7

The location of the breakaway point can be determined by solving the following equation for σ_b:

$$\sum_{i=1}^{n} \frac{1}{(\sigma_b + p_i)} = \sum_{i=1}^{m} \frac{1}{(\sigma_b + z_i)} \tag{13.8}$$

where $-p_i$ and $-z_i$ are the poles and zeros of GH, respectively. The solution of this equation requires

factorization of an $(n+m-1)$-order polynomial in σ_b. Consequently, the breakaway point can only be easily determined analytically for relatively simple GH. However, an approximate location can often be determined intuitively; then an iterative process can be used to solve the equation more exactly (see Problem 13.20). Computer programs for factorization of polynomials could also be applied.

EXAMPLE 13.5. To determine the breakaway points for $GH = K/s(s+1)(s+2)$, the following equation must be solved for σ_b:

$$\frac{1}{\sigma_b} + \frac{1}{\sigma_b+1} + \frac{1}{\sigma_b+2} = 0$$

$$(\sigma_b+1)(\sigma_b+2) + \sigma_b(\sigma_b+2) + \sigma_b(\sigma_b+1) = 0$$

which reduces to $3\sigma_b^2 + 6\sigma_b + 2 = 0$ whose roots are $\sigma_b = -0.423, -1.577$.

Applying the real axis rule of Section 13.5 for $K>0$ indicates that there are branches of the root-locus between 0 and -1 and between $-\infty$ and -2. Therefore the root at -0.423 is a breakaway point, as shown in Fig. 13-8. The value $\sigma_b = -1.577$ represents a breakaway on the root-locus for negative values of K since the portion of the real axis between -1 and -2 is on the root-locus for $K<0$.

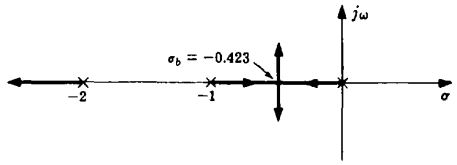


Fig. 13-8

13.8 DEPARTURE AND ARRIVAL ANGLES

The **departure angle** of the root-locus from a *complex pole* is given by

$$\theta_D = 180° + \arg GH' \tag{13.9}$$

where $\arg GH'$ is the phase angle of GH computed at the complex pole, but ignoring the contribution of that particular pole.

EXAMPLE 13.6. Consider the continuous system open-loop transfer function

$$GH = \frac{K(s+2)}{(s+1+j)(s+1-j)} \qquad K>0$$

The departure angle of the root-locus from the complex pole at $s = -1+j$ is determined as follows. The angle of GH for $s = -1+j$, ignoring the contribution of the pole at $s = -1+j$, is $-45°$. Therefore the departure angle is

$$\theta_D = 180° - 45° = 135°$$

and is illustrated in Fig. 13-9.

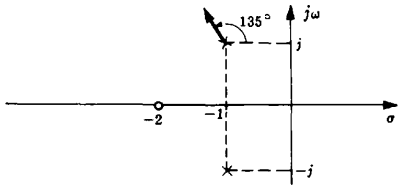


Fig. 13-9

The **angle of arrival** of the root-locus at a *complex zero* is given by

$$\theta_A = 180° - \arg GH'' \qquad (13.10)$$

where $\arg GH''$ is the phase angle of GH at the complex zero, ignoring the effect of that zero.

EXAMPLE 13.7. Consider the discrete-time system open-loop transfer function

$$\frac{K(z+j)(z-j)}{z(z+1)} \qquad K>0$$

The arrival angle of the root-locus for the complex zero at $z=j$ is $\theta_A = 180° - (-45°) = 225°$ as shown in Fig. 13-10.

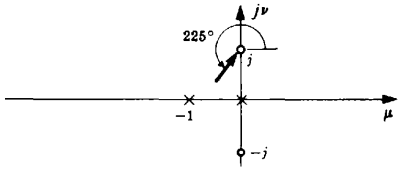


Fig. 13-10

13.9 CONSTRUCTION OF THE ROOT-LOCUS

A root-locus plot may be easily and accurately sketched using the construction rules of Sections 13.4 through 13.8. An efficient procedure is the following. First, determine the portions of the root-locus on the real axis. Second, compute the center and angles of the asymptotes and draw the asymptotes on the plot. Then determine the departure and arrival angles at complex poles and zeros (if any) and indicate them on the plot. Next, make a rough sketch of the branches of the root-locus so that each branch of the locus either terminates at a zero or approaches infinity along one of the asymptotes. The accuracy of this last step should of course improve with experience.

The accuracy of the plot may be improved by applying the angle criterion in the vicinity of the estimated branch locations. The rule of Section 13.7 can also be applied to determine the exact location of breakaway points.

The magnitude criterion of Section 13.3 is used to determine the values of K along the branches of the root-locus.

Since complex poles must occur in complex conjugate pairs (assuming real coefficients for the numerator and denominator polynomials of GH), the root-locus is symmetric about the real axis. Thus it is sufficient to plot only the upper half of the root-locus. However, it must be remembered that, in doing this, the lower halves of open-loop complex poles and zeros must be included when applying the magnitude and angle criteria.

Often, for analysis or design purposes, an accurate plot of the root-locus is required only in certain regions of the complex plane. In this case, the angle and magnitude criteria need only be applied in those regions of interest after a rough sketch has established the general shape of the plot. Of course, if a computer and appropriate software are available, plotting of even very complex root-loci can be a simple matter.

EXAMPLE 13.8. The root-locus for the closed-loop continuous system with open-loop transfer function

$$GH = \frac{K}{s(s+2)(s+4)} \qquad K>0$$

is constructed as follows. Applying the real axis rule of Section 13.5, the portions of the real axis between 0 and -2 and between -4 and $-\infty$ lie on the root-locus for $K>0$. The center of asymptotes is determined from Equation (13.6) to be $\sigma_c = -(2+4)/3 = -2$, and there are three asymptotes located at angles of $\beta = 60°$, 180°, and 300°.

Since two branches of the root-locus for $K>0$ come together on the real axis between 0 and -2, a breakaway point exists on that portion of the real axis. Hence the root-locus for $K>0$ may be sketched by estimating the location of the breakaway point and continuing the branches of the root-locus to the asymptotes, as shown in Fig. 13-11. To improve the accuracy of this plot, the exact location of the breakaway point is determined from Equation (13.8):

$$\frac{1}{\sigma_b} + \frac{1}{\sigma_b+2} + \frac{1}{\sigma_b+4} = 0$$

which simplifies to $3\sigma_b^2 + 12\sigma_b + 8 = 0$. The appropriate solution of this equation is $\sigma_b = -0.845$.

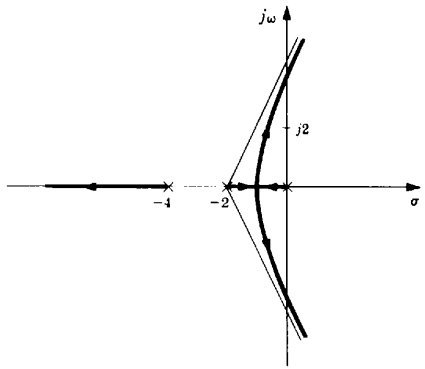


Fig. 13-11

The angle criterion is applied to points in the vicinity of the approximate root-locus to improve the accuracy of the location of the branches in the complex part of the s-plane; the magnitude criterion is used to determine the values of K along the root-locus. The resulting root-locus plot for $K>0$ is shown in Fig. 13-12.

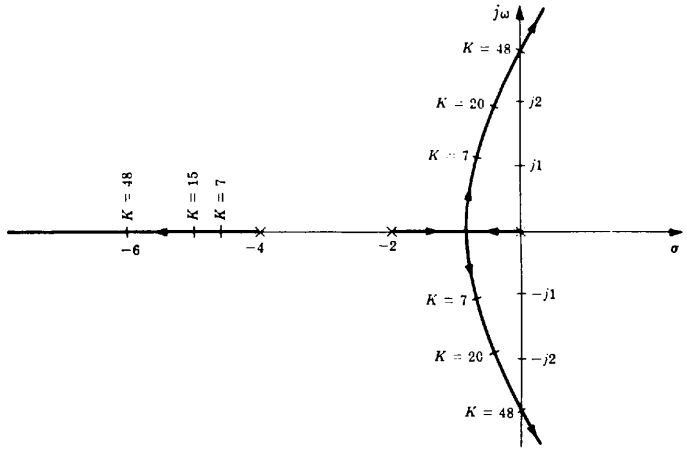


Fig. 13-12

The root-locus for $K<0$ is constructed in a similar manner. In this case, however, the portions of the real axis between 0 and ∞ and between -2 and -4 lie on the root-locus; the breakaway point is located at -3.155; and the asymptotes have angles of 0°, 120°, and 240°. The root-locus for $K<0$ is shown in Fig. 13-13.

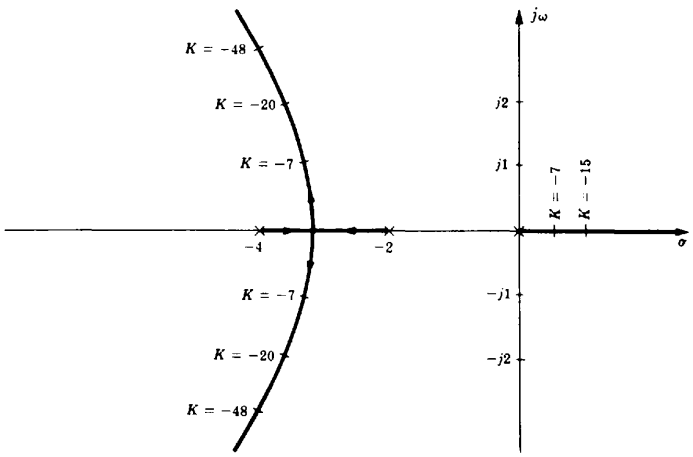


Fig. 13-13

13.10 THE CLOSED-LOOP TRANSFER FUNCTION AND THE TIME-DOMAIN RESPONSE

The closed-loop transfer function C/R is easily determined from the root-locus plot for a specified value of open-loop gain factor K. From this, the time-domain response $c(t)$ may be determined for a given Laplace transformable input $r(t)$ for continuous systems by inversion of $C(s)$. For discrete systems, $c(k)$ can be similarly determined by inversion of $C(z)$.

Consider the closed-loop transfer function C/R for the canonical *unity* (*negative*) *feedback* system

$$\frac{C}{R} = \frac{G}{1+G} \tag{13.11}$$

Open-loop transfer functions which are rational algebraic expressions can be written (for continuous systems) as

$$G = \frac{KN}{D} = \frac{K(s+z_1)(s+z_2)\cdots(s+z_m)}{(s+p_1)(s+p_2)\cdots(s+p_n)} \tag{13.12}$$

G has the same form for discrete-time systems, with z replacing s in Equation (*13.12*). In Equation (*13.12*), $-z_i$ are the zeros, $-p_i$ are the poles of G, $m \le n$, and N and D are polynomials whose roots are $-z_i$ and $-p_i$, respectively. Then

$$\frac{C}{R} = \frac{KN}{D+KN} \tag{13.13}$$

and it is clear that C/R and G have the same zeros but not the same poles (unless $K = 0$). Hence

$$\frac{C}{R} = \frac{K(s+z_1)(s+z_2)\cdots(s+z_m)}{(s+\alpha_1)(s+\alpha_2)\cdots(s+\alpha_n)}$$

where $-\alpha_i$ denote the n closed-loop poles. The location of these poles is by definition determined directly from the root-locus plot for a specified value of open-loop gain K.

EXAMPLE 13.9. Consider the continuous system whose open-loop transfer function is

$$G = \frac{K(s+2)}{(s+1)^2} \qquad K > 0$$

The root-locus plot is given in Fig. 13-14.

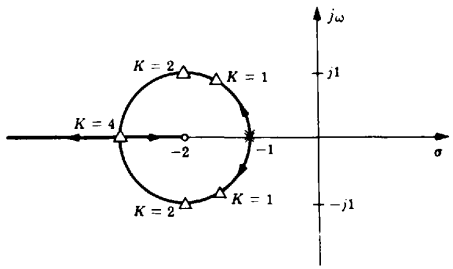


Fig. 13-14

Several values of gain factor K are shown at points on the loci denoted by *small triangles*. These points are the *closed-loop poles* corresponding to the specified values of K. For $K = 2$, the closed-loop poles are $-\alpha_1 = -2 + j$ and $-\alpha_2 = -2 - j$. Therefore

$$\frac{C}{R} = \frac{2(s+2)}{(s+2+j)(s+2-j)}$$

When the system is not unity feedback, then

$$\frac{C}{R} = \frac{G}{1+GH} \tag{13.14}$$

and

$$GH = \frac{KN}{D} \tag{13.15}$$

The closed-loop poles may be determined directly from the root-locus for a given K, but the closed-loop zeros are not equal to the open-loop zeros. The open-loop zeros must be computed separately by clearing fractions in Equation (*13.14*).

EXAMPLE 13.10. Consider the continuous system described by

$$G = \frac{K(s+2)}{s+1} \qquad H = \frac{1}{s+1} \qquad GH = \frac{K(s+2)}{(s+1)^2} \qquad K > 0$$

and

$$\frac{C}{R} = \frac{K(s+1)(s+2)}{(s+1)^2 + K(s+2)} = \frac{K(s+1)(s+2)}{(s+\alpha_1)(s+\alpha_2)}$$

The root-locus plot for this example is the same as that for Example 13.9. Hence for $K = 2$, $\alpha_1 = 2 + j$ and $\alpha_2 = 2 - j$. Thus

$$\frac{C}{R} = \frac{2(s+1)(s+2)}{(s+2+j)(s+2-j)}$$

EXAMPLE 13.11. For the discrete-time system with $GH(z) = K/z(z-1)$, the root-locus for $K > 0$ is shown in Fig. 13-15. For $K = 0.25$, the roots are at $z = 0.5$ and the closed-loop transfer function is

$$\frac{C}{R} = \frac{0.25}{(z-0.5)^2}$$

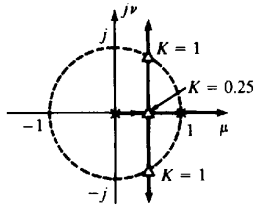


Fig. 13-15

13.11 GAIN AND PHASE MARGINS FROM THE ROOT-LOCUS

The **gain margin** is the factor by which the gain factor K can be multiplied before the closed-loop system becomes unstable. It can be determined from the root-locus using the following formula:

$$\text{gain margin} = \frac{\text{value of } K \text{ at the stability boundary}}{\text{design value of } K} \qquad (13.16)$$

where the stability boundary is the $j\omega$-axis in the s-plane, or the unit circle in the z-plane. If the root-locus does not cross the stability boundary, the gain margin is infinite.

EXAMPLE 13.12. Consider the continuous system in Fig. 13-16. The design value for the gain factor is 8, producing the closed-loop poles (denoted by small triangles) shown in the root-locus of Fig. 13-17. The gain factor at the $j\omega$-axis crossing is 64; hence the gain margin for this system is $64/8 = 8$.

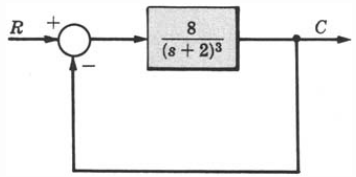


Fig. 13-16

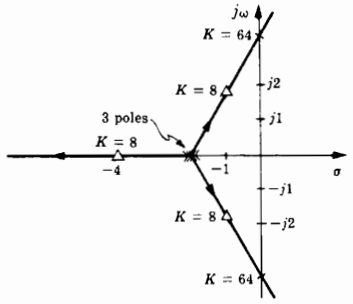


Fig. 13-17

EXAMPLE 13.13. The root-locus for the discrete-time system of Example 13.11 crosses the stability boundary (unit circle) for $K = 1$. For a design value of $K = 0.25$, the gain margin is $1/0.25 = 4$.

The **phase margin** can also be determined from the root-locus. In this case it is necessary to find the point ω_1 on the stability boundary for which $|GH| = 1$ for the design value of K; that is,

$$|D(\omega_1)/N(\omega_1)| = K_{\text{design}}$$

It is usually necessary to use a trial-and-error procedure to locate ω_1. The phase margin is then computed from $\arg GH(\omega_1)$ as

$$\phi_{PM} = [180° + \arg GH(\omega_1)] \text{ degrees} \qquad (13.17)$$

EXAMPLE 13.14. For the system of Example 13.12, $|GH(\omega_1)| = |8/(j\omega_1 + 2)^3| \equiv 1$ when $\omega_1 = 0$; the phase angle of $GH(0)$ is 0°. The phase margin is therefore 180°.

EXAMPLE 13.15. For the continuous system of Fig. 13-18, the root-locus is shown in Fig. 13-19. The point on the $j\omega$-axis for which $|GH(\omega_1)| = |24/j\omega_1(j\omega_1 + 4)^2| \equiv 1$ is at $\omega_1 = 1.35$; the angle of $GH(1.35)$ is $-129.6°$. Therefore the phase margin is $\phi_{PM} = 180° - 129.6° = 50.4°$.

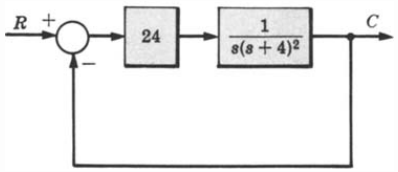


Fig. 13-18

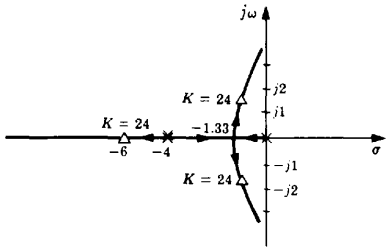


Fig. 13-19

13.12 DAMPING RATIO FROM THE ROOT-LOCUS FOR CONTINUOUS SYSTEMS

The gain factor K required to give a specified damping ratio ζ (or vice versa) for the second-order continuous system

$$GH = \frac{K}{(s + p_1)(s + p_2)} \qquad K, p_1, p_2 > 0$$

is easily determined from the root-locus. Simply draw a line from the origin at an angle of plus or minus θ with the negative real axis, where

$$\theta = \cos^{-1}\zeta \qquad (13.18)$$

(See Section 4.13.) The gain factor at the point of intersection with the root-locus is the required value of K. This procedure can be applied to any pair of complex conjugate poles, for systems of second or higher order. For higher-order systems, the damping ratio determined by this procedure for a *specific pair* of complex poles does not necessarily determine the damping (predominant time constant) of the system.

EXAMPLE 13.16. Consider the third-order system of Example 13.15. The damping ratio ζ of the *complex poles* for $K = 24$ is easily determined by drawing a line from the origin to the point on the root-locus where $K = 24$, as shown in Fig. 13-20. The angle θ is measured as 60°; hence

$$\zeta = \cos\theta = 0.5$$

This value of ζ is a good approximation for the damping of the third-order system with $K = 24$ because the complex poles dominate the response.

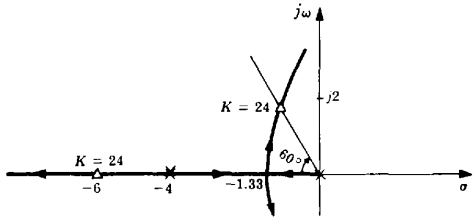


Fig. 13-20

Solved Problems

VARIATION OF SYSTEM CLOSED-LOOP POLES

13.1. Determine the closed-loop transfer function and the characteristic equation of the unity negative feedback control system whose open-loop transfer function is $G = K(s + 2)/(s + 1)(s + 4)$.

The closed-loop transfer function is

$$\frac{C}{R} = \frac{G}{1+G} = \frac{K(s+2)}{(s+1)(s+4)+K(s+2)}$$

The characteristic equation is obtained by setting the denominator polynomial equal to zero:

$$(s+1)(s+4)+K(s+2)=0$$

13.2. How would the closed-loop poles of the system of Problem 13.1 be determined for $K=2$ from its root-locus plot?

The root-locus is a plot of the closed-loop poles of the feedback system as a function of K. Therefore the closed-loop poles for $K=2$ are determined by the points on the root-locus which correspond to $K=2$ (one point on each branch of the locus).

13.3. How can a root-locus be employed to factor the polynomial $s^2+6s+18$?

Since the root-locus is a plot of the roots of the characteristic equation of a system, Equation (*13.1*), as a function of its open-loop gain factor, the roots of the above polynomial can be determined from the root-locus of any system whose characteristic polynomial is equivalent to it for some value of K. For example, the root-locus for $GH = K/s(s+6)$ factors the characteristic polynomial s^2+6s+K. For $K=18$ this polynomial is equivalent to the one we desire to factor. Thus the desired roots are located on this root-locus at the points corresponding to $K=18$.

Note that other forms for GH could be chosen, such as $GH = K/(s+2)(s+4)$ whose closed-loop characteristic polynomial corresponds to the one we wish to factor, but now for $K=10$.

ANGLE AND MAGNITUDE CRITERIA

13.4. Show that the point $p_1 = -0.5$ satisfies the angle criterion, Equation (*13.4*), and the magnitude criterion, Equation (*13.5*), when $K=1.5$ in the open-loop transfer function of Example 13.1.

$$\arg GH(p_1) = \arg\frac{K(p_1+1)}{p_1(p_1+2)} = \arg\frac{1.5(0.5)}{-0.5(1.5)} = 180^\circ \qquad |GH(p_1)| = \left|\frac{1.5(0.5)}{-0.5(1.5)}\right| = 1$$

or

$$\left|\frac{D(p_1)}{N(p_1)}\right| = \left|\frac{-0.5(1.5)}{0.5}\right| = 1.5 = K$$

Thus as illustrated on the root-locus plot of Example 13.1, the point $p_1 = -0.5$ is on the root-locus and is a closed-loop pole for $K=1.5$.

13.5. Determine the angle and magnitude of $GH(j2)$ for $GH = K/s(s+2)^2$. What value of K satisfies $|GH(j2)| = 1$?

$$GH(j2) = \frac{K}{j2(j2+2)^2} \qquad \arg GH(j2) = \begin{cases} -180^\circ & \text{for } K>0 \\ 0^\circ & \text{for } K<0 \end{cases} \qquad |GH(j2)| = \frac{|K|}{2(8)} = \frac{|K|}{16}$$

and for $|GH(j2)| = 1$ it is necessary that $|K| = 16$.

13.6. Illustrate the graphical composition of $\arg GH(j2)$ and $|GH(j2)|$ in Problem 13.5.

$$\arg GH(j2) = -90^\circ - 45^\circ - 45^\circ = -180^\circ \qquad |GH(j2)| = \frac{|K|}{2(2\sqrt{2})^2} = \frac{|K|}{16}$$

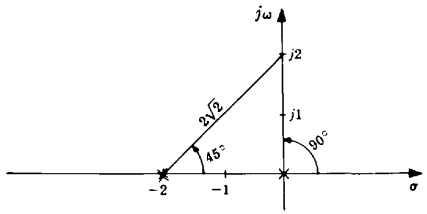


Fig. 13-21

13.7. Show that the point $p_1 = -1+j\sqrt{3}$ is on the root-locus for

$$GH(s) = \frac{K}{(s+1)(s+2)(s+4)} \qquad K>0$$

and determine K at this point.

$$\arg\frac{N(p_1)}{D(p_1)} = \arg\frac{1}{j\sqrt{3}(1+j\sqrt{3})(3+j\sqrt{3})} = -90^\circ - 60^\circ - 30^\circ = -180^\circ$$

The angle criterion, Equation (*13.4b*), is thus satisfied for $K>0$ and the point $p_1 = -1+j\sqrt{3}$ is on the root-locus. From Equation (*13.5*),

$$K = \left|\frac{j\sqrt{3}(1+j\sqrt{3})(3+j\sqrt{3})}{1}\right| = \sqrt{3(4)12} = 12$$

NUMBER OF LOCI

13.8. Why must the number of loci equal the number of open-loop poles for $m \le n$?

Each branch of the root-locus represents the locus of one closed-loop pole. Consequently there must be as many branches or loci as there are closed-loop poles. Since the number of closed-loop poles is equal to the number of open-loop poles for $m \le n$, the number of loci must equal the number of open-loop poles.

13.9. How many loci are in the root-locus for

$$GH(z) = \frac{K(z+\frac{1}{3})(z+\frac{1}{2})}{z(z+\frac{1}{2}+j/2)(z-\frac{1}{2}-j/2)}$$

Since the number of open-loop poles is three, there are three loci in the root-locus plot.

REAL AXIS LOCI

13.10. Prove the real axis loci rules.

For any point on the real axis, the angle contributed to $\arg GH$ by any real axis pole or zero is either 0° or 180°, depending on whether or not the point is to the right or to the left of the pole or zero. The total angle contributed to $\arg GH(s)$ by a pair of complex poles or zeros is zero because

$$\arg(s+\sigma_1+j\omega_1) + \arg(s+\sigma_1-j\omega_1) = 0$$

for all real values of s. Thus $\arg GH(s)$ for real values of s ($s=\sigma$) may be written as

$$\arg GH(\sigma) = 180n_r + \arg K$$

where n_r is the total number of finite poles and zeros to the right of σ. In order to satisfy the angle criterion, n_r must be odd for positive K and even for negative K. Thus for $K>0$, points of the root-locus on the real axis lie to the left of an odd number of finite poles and zeros; and for $K<0$, points of the root-locus on the real axis lie to the left of an even number of finite poles and zeros.

13.11. Determine which parts of the real axis are on the root-locus of

$$GH = \frac{K(s+2)}{(s+1)(s+3+j)(s+3-j)} \qquad K>0$$

The points on the real axis which lie to the left of an odd number of finite poles and zeros are only those points between -1 and -2. Therefore by the rule for $K>0$, only the portion of real axis between -1 and -2 lies on the root-locus.

13.12. Which parts of the real axis are on the root-locus for

$$GH = \frac{K}{s(s+1)^2(s+2)} \qquad K>0$$

Points on the real axis between 0 and -1 and between -1 and -2 lie to the left of an odd number of poles and zeros and therefore are on the root-locus for $K>0$.

ASYMPTOTES

13.13. Prove that the angles of the asymptotes are given by

$$\beta = \begin{cases} \dfrac{(2l+1)180}{n-m} \text{ degrees} & \text{for } K>0 \\ \dfrac{(2l)180}{n-m} \text{ degrees} & \text{for } K<0 \end{cases} \qquad (13.7)$$

For points s far from the origin in the s-plane, the angle contributed to $\arg GH$ by each of m zeros is

$$\arg(s+z_i)\big|_{|s|\gg|z_i|} \cong \arg(s)$$

Similarly, the angle contributed to $\arg GH$ by each of n poles is approximately equal to $-\arg(s)$. Therefore

$$\arg\left[\frac{N(s)}{D(s)}\right] \cong -(n-m)\cdot\arg(s) = -(n-m)\beta$$

where $\beta \equiv \arg(s)$. In order for s to be on the root-locus the angle criterion, Equation (*13.4b*), must be satisfied. Thus

$$\arg\left[\frac{N(s_1)}{D(s_1)}\right] = -(n-m)\beta = \begin{cases} (2l+1)\pi & \text{for } K>0 \\ (2l)\pi & \text{for } K<0 \end{cases}$$

and, since $\pm\pi$ radians ($\pm 180°$) are the same angle in the s-plane, then

$$\beta = \begin{cases} \dfrac{(2l+1)180}{n-m} \text{ degrees} & \text{for } K>0 \\ \dfrac{(2l)180}{n-m} \text{ degrees} & \text{for } K<0 \end{cases}$$

The proof is similar for the z-plane.

13.14. Show that the center of asymptotes is given by

$$\sigma_c = -\frac{\sum_{i=1}^{n} p_i - \sum_{i=1}^{m} z_i}{n-m} \qquad (13.6)$$

The points on the root-locus satisfy the characteristic equation $D+KN=0$, or

$$s^n + b_{n-1}s^{n-1} + \cdots + b_0 + K\left(s^m + a_{m-1}s^{m-1} + \cdots + a_0\right) = 0$$

Dividing by the numerator polynomial $N(s)$, this becomes

$$s^{n-m} + (b_{n-1} - a_{m-1})s^{n-m-1} + \cdots + K = 0$$

(same for the z-plane, with z replacing s). When the first coefficient of a polynomial is unity, the second coefficient is equal to minus the sum of the roots (see Problem 5.26). Thus from $D(s)=0$, $b_{n-1} = \sum_{i=1}^{n} p_i$. From $N(s)=0$, $a_{m-1} = \sum_{i=1}^{m} z_i$; and $-(b_{n-1} - a_{m-1})$ is equal to the sum of $n-m$ roots of the characteristic equation.

Now for large values of K and correspondingly large distances from the origin these $n-m$ roots approach the straight-line asymptotes and, along the asymptotes, the sum of the $n-m$ roots is equal to $-(b_{n-1} - a_{m-1})$. Since $b_{n-1} - a_{m-1}$ is a real number, the asymptotes must intersect at a point on the real axis. The center of asymptotes is therefore given by the point on the real axis where $n-m$ equal roots add up to $-(b_{n-1} - a_{m-1})$. Thus

$$\sigma_c = -\frac{b_{n-1} - a_{m-1}}{n-m} = -\frac{\sum_{i=1}^{n} p_i - \sum_{i=1}^{m} z_i}{n-m}$$

For a more detailed proof, see reference [6].

13.15. Find the angles and center of, and sketch the asymptotes for

$$GH = \frac{K(s+2)}{(s+1)(s+3+j)(s+3-j)(s+4)} \qquad K>0$$

The center of asymptotes is

$$\sigma_c = -\frac{1+3+j+3-j+4-2}{4-1} = -3$$

There are three asymptotes located at angles of $\beta = 60°$, $180°$, and $300°$ as shown in Fig. 13-22.

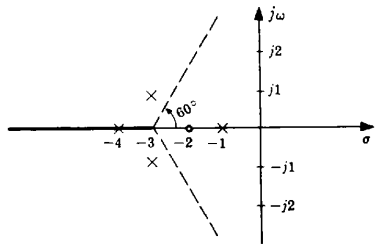


Fig. 13-22

13.16. Sketch the asymptotes for $K>0$ and $K<0$ for

$$GH = \frac{K}{s(s+2)(s+1+j)(s+1-j)}$$

The center of asymptotes is $\sigma_c = -(0+2+1+j+1-j)/4 = -1$.
For $K>0$, the angles of the asymptotes are $\beta = 45°$, $135°$, $225°$, and $315°$ as shown in Fig. 13-23.
For $K<0$, the angles of the asymptotes are $\beta = 0°$, $90°$, $180°$, and $270°$ as shown in Fig. 13-24.

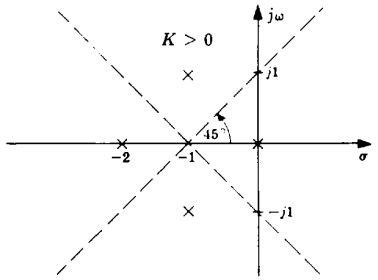


Fig. 13-23

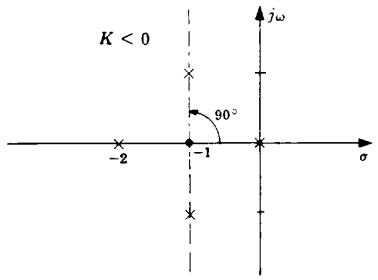


Fig. 13-24

BREAKAWAY POINTS

13.17. Show that a breakaway point σ_b satisfies

$$\sum_{i=1}^{n} \frac{1}{(\sigma_b + p_i)} = \sum_{i=1}^{m} \frac{1}{(\sigma_b + z_i)} \qquad (13.8)$$

A breakaway point is a point on the real axis where the gain factor K along the real axis portion of the root-locus is a maximum for poles leaving the real axis, or a minimum for poles coming onto the real axis, (see Section 13.2). The gain factor along the root-locus is given by

$$|K| = \left|\frac{D}{N}\right| \qquad (13.5)$$

On the real axis, $s = \sigma$ (or $z = \mu$) and the magnitude signs may be dropped because $D(\sigma)$ and $N(\sigma)$ are both real. Then

$$K = \frac{D(\sigma)}{N(\sigma)}$$

To find the value of σ for which K is a maximum or minimum, the derivative of K with respect to σ is set equal to zero:

$$\frac{dK}{d\sigma} = \frac{d}{d\sigma}\left[\frac{(\sigma + p_1)\cdots(\sigma + p_n)}{(\sigma + z_1)\cdots(\sigma + z_m)}\right] = 0$$

By repeated differentiation and factorization, this can be written as

$$\frac{dK}{d\sigma} = \sum_{i=1}^{n} \frac{1}{(\sigma + p_i)}\left[\frac{D(\sigma)}{N(\sigma)}\right] - \sum_{i=1}^{m} \frac{1}{(\sigma + z_i)}\left[\frac{D(\sigma)}{N(\sigma)}\right] = 0$$

Finally, dividing both sides by $D(\sigma)/N(\sigma)$ yields the required result.

13.18. Determine the breakaway point for $GH = K/s(s+3)^2$.

The breakaway point satisfies

$$\frac{1}{\sigma_b} + \frac{1}{\sigma_b + 3} + \frac{1}{\sigma_b + 3} = 0$$

from which $\sigma_b = -1$.

13.19. Find the breakaway point for

$$GH = \frac{K(s+2)}{(s+1+j\sqrt{3})(s+1-j\sqrt{3})}$$

From Equation (*13.8*),

$$\frac{1}{\sigma_b + 1 + j\sqrt{3}} + \frac{1}{\sigma_b + 1 - j\sqrt{3}} = \frac{1}{\sigma_b + 2}$$

which gives $\sigma_b^2 + 4\sigma_b = 0$. This equation has the solution $\sigma_b = 0$ and $\sigma_b = -4$; $\sigma_b = -4$ is a breakaway point for $K > 0$ and $\sigma_b = 0$ is a breakaway point for $K < 0$, as shown in Fig. 13-25.

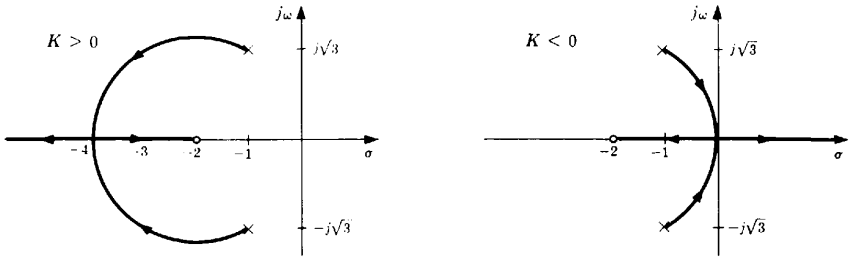


Fig. 13-25

13.20. Find the breakaway point between 0 and -1 for

$$GH = \frac{K}{s(s+1)(s+3)(s+4)}$$

The breakaway point must satisfy

$$\frac{1}{\sigma_b} + \frac{1}{(\sigma_b + 1)} + \frac{1}{(\sigma_b + 3)} + \frac{1}{(\sigma_b + 4)} = 0$$

If this equation were simplified, a third-order polynomial would be obtained. To avoid solving a third-order polynomial, the following procedure may be used. As a first guess, assume $\sigma_b = -0.5$ and use this value in the two terms for the poles furthest from the breakaway point. Then

$$\frac{1}{\sigma_b} + \frac{1}{\sigma_b + 1} + \frac{1}{2.5} + \frac{1}{3.5} = 0$$

which simplifies to $\sigma_b^2 + 3.92\sigma_b + 1.46 = 0$ and has the root $\sigma_b = -0.43$ between 0 and -1. This value is used to obtain a better approximation as follows:

$$\frac{1}{\sigma_b} + \frac{1}{\sigma_b + 1} + \frac{1}{2.57} + \frac{1}{3.57} = 0 \qquad \sigma_b^2 + 3.99\sigma_b + 1.496 = 0 \qquad \sigma_b = -0.424$$

The second computation did not result in a value much different from the first. A reasonable first guess can often result in a fairly accurate approximation with only one computation.

DEPARTURE AND ARRIVAL ANGLES

13.21. Show that the departure angle of the root-locus from a complex pole is given by

$$\theta_D = 180° + \arg GH' \qquad (13.9)$$

Consider a circle of infinitesimally small radius around the complex pole. Clearly, the phase angle $\arg GH'$ of GH, neglecting the contribution of the complex pole, is constant around this circle. If θ_D represents the departure angle, the total phase angle of GH at the point on the circle where the root-locus crosses it is

$$\arg GH = \arg GH' - \theta_D$$

since $-\theta_D$ is the phase angle contributed to $\arg GH$ by the complex pole. In order to satisfy the angle criterion, $\arg GH = \arg GH' - \theta_D = 180°$ or $\theta_D = 180° + \arg GH'$ since $+180°$ and $-180°$ are equivalent.

13.22. Determine the relationship between the departure angle from a complex pole for $K > 0$ with that for $K < 0$.

Since $\arg GH'$ changes by 180° if K changes from a positive number to a negative one, the departure angle for $K < 0$ is 180° different from the departure angle for $K > 0$.

13.23. Show that the arrival angle at a complex zero satisfies

$$\theta_A = 180° - \arg GH'' \qquad (13.10)$$

In the same manner as in the solution to Problem 13.21, the phase angle of GH in the vicinity of the complex zero is given by $\arg GH = \arg GH'' + \theta_A$ since θ_A is the phase angle contributed to $\arg GH$ by the complex zero. Then applying the angle criterion yields $\theta_A = 180° - \arg GH''$.

13.24. Graphically determine $\arg GH'$ and compute the departure angle of the root-locus from the complex pole at $s = -2 + j$ for

$$GH = \frac{K}{(s+1)(s+2-j)(s+2+j)} \qquad K > 0$$

From Fig. 13-26, $\arg GH' = -135° - 90° = -225°$; and $\theta_D = 180° - 225° = -45°$ as shown in Fig. 13-27.

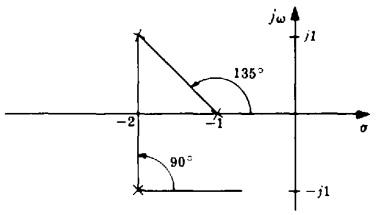


Fig. 13-26

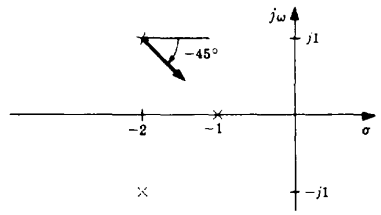


Fig. 13-27

13.25. Determine the departure angles from the complex poles and the arrival angles at the complex zeros for the open-loop transfer function

$$GH = \frac{K(s+1+j)(s+1-j)}{s(s+2j)(s-2j)} \qquad K > 0$$

For the complex pole at $s = 2j$,

$$\arg GH' = 45° + 71.6° - 90° - 90° = -63.4° \quad \text{and} \quad \theta_D = 180° - 63.4° = 116.6°$$

Since the root-locus is symmetric about the real axis, the departure angle from the pole at $s = -2j$ is $-116.6°$. For the complex zero $s = -1 + j$,

$$\arg GH'' = 90° - 108.4° - 135° - 225° = -18.4° \quad \text{and} \quad \theta_A = 180° - (-18.4°) = 198.4°$$

Thus the arrival angle at the complex zero $s = -1 - j$ is $\theta_A = -198.4°$.

CONSTRUCTION OF THE ROOT-LOCUS

13.26. Construct the root-locus for

$$GH = \frac{K}{(s+1)(s+2-j)(s+2+j)} \qquad K > 0$$

The real axis from -1 to $-\infty$ is on the root-locus. The center of asymptotes is at

$$\sigma_c = \frac{-1-2+j-2-j}{3} = -1.67$$

There are three asymptotes ($n - m = 3$), located at angles of 60°, 180°, and 300°. The departure angle from the complex pole at $s = -2 + j$ computed in Problem 13.24 is $-45°$. A sketch of the resulting root-locus is shown in Fig. 13-28. An accurate root-locus plot is obtained by checking the angle criterion at points along the sketched branches, adjusting the location of the branches if necessary, and then applying the magnitude criterion to determine the values of K at selected points along the branches. The completed root-locus is shown in Fig. 13-29.

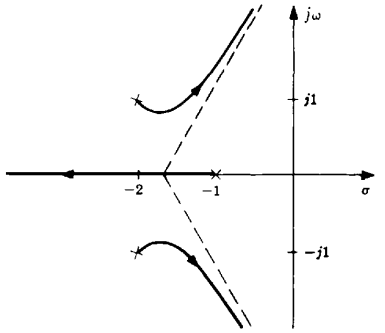


Fig. 13-28

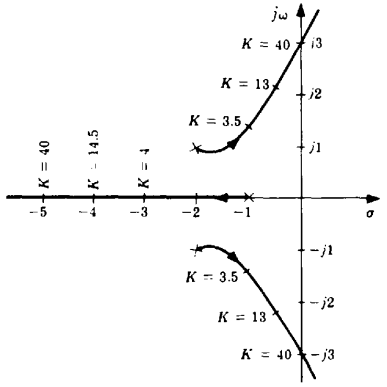


Fig. 13-29

13.27. Sketch the branches of the root-locus for the transfer function

$$GH = \frac{K(s+2)}{(s+1)(s+3+j)(s+3-j)} \qquad K > 0$$

The real axis between -1 and -2 is on the root-locus (Problem 13.11). There are two asymptotes with angles of 90° and 270°. The center of asymptotes is easily computed as $\sigma_c = -2.5$ and the departure angle from the complex pole at $s = -3 + j$ as 72°. By symmetry, the departure angle from the pole at $-3 - j$ is $-72°$. The branches of the root-locus may therefore be sketched as shown in Fig. 13-30.

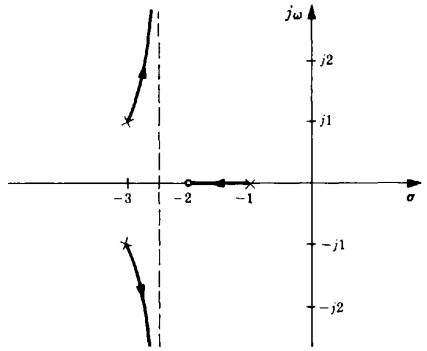


Fig. 13-30

13.28. Construct the root-locus for $K > 0$ and $K < 0$ for the transfer function

$$GH = \frac{K}{s(s+1)(s+3)(s+4)}$$

For this transfer function the center of asymptotes is simply $\sigma_c = -2$; and $n - m = 4$. Therefore for $K > 0$ the asymptotes have angles of 45°, 135°, 225°, and 315°. The real axis sections between 0 and -1 and between -3 and -4 lie on the root-locus for $K > 0$ and it was determined in Problem 13.20 that a breakaway point is located at $\sigma_b = -0.424$. From the symmetry of the pole locations, another breakaway point is located at -3.576. This can be verified by substituting this value into the relation for the breakaway point, Equation (*13.8*). The completed root-locus for $K > 0$ is shown in Fig. 13-31.

For $K < 0$, the asymptotes have angles of 0°, 90°, 180°, and 270°. In this case the real axis portions between ∞ and 0, between -1 and -3, and between -4 and $-\infty$ are on the root-locus. There is only one breakaway point, located at -2. The completed root-locus for $K < 0$ is shown in Fig. 13-32.

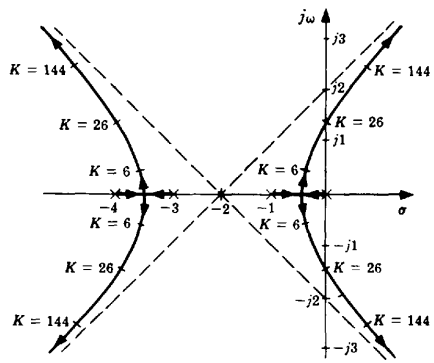


Fig. 13-31

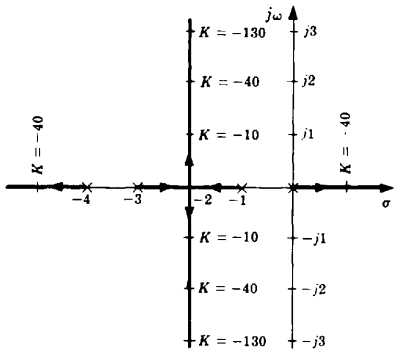


Fig. 13-32

13.29. Construct the root-locus for $K > 0$ for the discrete system transfer function

$$GH(z) = \frac{K(z-0.5)}{(z-1)^2}$$

This root-locus has two loci and one asymptote. The root-locus lies on the real axis for $z < 0.5$. The breakaway points are at $z = 0$ and $z = 1$. The completed root-locus is shown in Fig. 13-33.

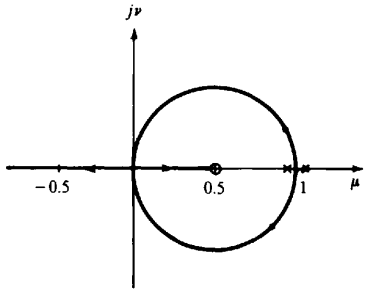


Fig. 13-33

13.30. Construct the root-locus for $K > 0$ for the discrete system transfer function

$$GH(z) = \frac{K}{(z+0.5)(z-1.5)}$$

This root-locus has two branches and two asymptotes. The breakaway point and the center of asymptotes are at $z = 0.5$. The root-locus is shown in Fig. 13-34.

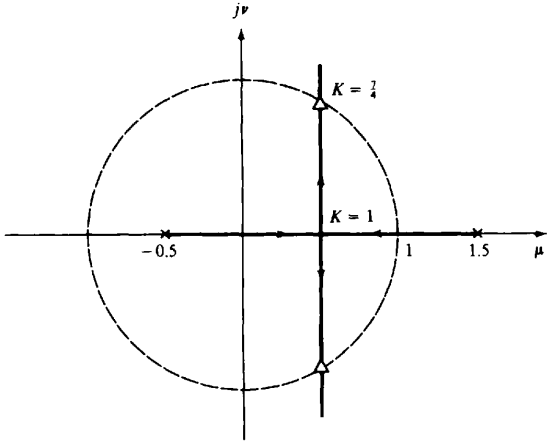


Fig. 13-34

13.31. Construct the root-locus for $K > 0$ for the discrete-time system with $H = 1$ and forward transfer function

$$G(z) = \frac{K(z+\frac{1}{3})(z+1)}{z(z+\frac{1}{2})(z-1)}$$

The system has one more pole than zero, so the root-locus has only one asymptote, along the negative real axis. The root-locus is on the real axis between 0 and 1, between $-\frac{1}{3}$ and $-\frac{1}{2}$, and to the left of -1. Breakaway points are located between 0 and 1 and to the left of -1. By trial and error (or computer solution), breakaway points are found at $z = 0.383$ and $z = -2.22$.

The root-locus is an ellipse between the breakaway points at $z = 0.383$ and $z = -2.22$. The point on the $j\nu$-axis, where $\arg G(z) = -180°$ is found by trial and error to be $z = j0.85$. Similarly, the point on the line $z = -1 + j\nu$, where $\arg G(z) = -180°$ is $z = -1 + j1.26$. The root-locus is drawn in Fig. 13-35. The gain factor along the root-locus is determined graphically from the pole-zero map or analytically by evaluating $G(z)$.

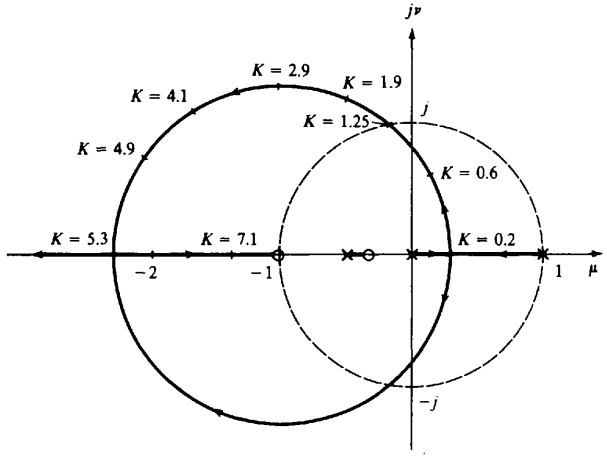


Fig. 13-35

THE CLOSED-LOOP TRANSFER FUNCTION AND THE TIME-DOMAIN RESPONSE

13.32. Determine the closed-loop transfer function of the continuous system of Example 13.8 for $K = 48$, given the following transfer functions for H: (*a*) $H = 1$, (*b*) $H = 4/(s+1)$, (*c*) $H = (s+1)/(s+2)$.

From the root-locus plot of Example 13.8, the closed-loop poles for $K = 48$ are located at $s = -6$, $j2.83$, and $-j2.83$. For $H = 1$,

$$G = \frac{48}{s(s+2)(s+4)} \quad \text{and} \quad \frac{C}{R} = \frac{GH}{1+GH} = \frac{48}{(s+6)(s-j2.83)(s+j2.83)}$$

For $H = 4/(s+1)$,

$$G = \frac{12(s+1)}{s(s+2)(s+4)} \quad \text{and} \quad \frac{C}{R} = \frac{1}{H}\left(\frac{GH}{1+GH}\right) = \frac{12(s+1)}{(s+6)(s-j2.83)(s+j2.83)}$$

For $H = (s+1)/(s+2)$,

$$G = \frac{48}{s(s+1)(s+4)} \quad \text{and} \quad \frac{C}{R} = \frac{48(s+2)}{(s+1)(s+6)(s-j2.83)(s+j2.83)}$$

Note that in this last case there are four closed-loop poles, while GH has only three poles. This is due to the cancellation of a pole of G by a zero of H.

13.33. Determine the unit step response of the system of Example 13.1 with $K = 1.5$.

The closed-loop transfer function of this system is

$$\frac{C}{R} = \frac{1.5(s+1)}{(s+0.5)(s+3)}$$

For $R = 1/s$,

$$C = \frac{1.5(s+1)}{s(s+0.5)(s+3)} = \frac{1}{s} + \frac{-0.6}{s+0.5} + \frac{-0.4}{s+3}$$

and the unit step response is $\mathcal{L}^{-1}[C(s)] = c(t) = 1 - 0.6e^{-0.5t} - 0.4e^{-3t}$.

13.34. Determine the relationship between the closed-loop zeros and the poles and zeros of G and H, assuming there are no cancellations.

Let $G = N_1/D_1$ and $H = N_2/D_2$, where N_1 and D_1 are numerator (zeros) and denominator (poles) polynomials of G, and N_2 and D_2 are the numerator and denominator polynomials of H. Then

$$\frac{C}{R} = \frac{G}{1+GH} = \frac{N_1 D_2}{D_1 D_2 + N_1 N_2}$$

Thus the closed-loop zeros are equal to the zeros of G and the poles of H.

GAIN AND PHASE MARGINS

13.35. Find the gain margin of the system of Example 13.8 for $K = 6$.

The gain factor at the $j\omega$-axis crossover is 48, as shown in Fig. 13-12. Hence the gain margin is $48/6 = 8$.

13.36. Show how a Routh table (Section 5.3) can be used to determine the frequency and the gain at the $j\omega$-axis crossover.

In Section 5.3 it was pointed out that a row of zeros in the s^1 row of the Routh table indicates that the polynomial has a pair of roots which satisfy the auxiliary equation $As^2 + B = 0$, where A and B are the first and second elements of the s^2 row. If A and B have the same sign, the roots of the auxiliary equation are imaginary (on the $j\omega$-axis). Thus if a Routh table is constructed for the characteristic equation of a system, the values of K and ω corresponding to $j\omega$-axis crossovers can be determined. For example, consider the system with the open-loop transfer function

$$GH = \frac{K}{s(s+2)^2}$$

The characteristic equation for this system is

$$s^3 + 4s^2 + 4s + K = 0$$

The Routh table for the characteristic polynomial is

s^3	1	4
s^2	4	K
s^1	$(16-K)/4$	
s^0	K	

The s^1 row is zero for $K = 16$. The auxiliary equation then becomes

$$4s^2 + 16 = 0$$

Thus for $K = 16$ the characteristic equation has solutions (closed-loop poles) at $s = \pm j2$, and the root-locus crosses the $j\omega$-axis at $j2$.

13.37. Determine the phase margin for the system of Example 13.8 (Figure 13-12) for $K = 6$.

First, the point on the $j\omega$-axis for which $|GH(j\omega)| = 1$ is found by trial and error to be $j0.7$. Then $\arg GH(j0.7)$ is computed as $-120°$. Hence the phase margin is $180° - 120° = 60°$.

13.38. Is it necessary to construct the entire root-locus in order to determine the gain and phase margins of a system?

No. Only one point on the root-locus is required to determine the gain margin. This point, at ω_n, where the root-locus crosses the stability boundary, can be determined by trial and error or by the use of a Routh table as described in Problem 13.36. To determine the phase margin, it is only necessary to determine the point on the stability boundary where $|GH(j\omega)| = 1$. Although the entire root-locus plot is not necessary, it can often be helpful, especially in the case of multiple stability boundary crossings.

DAMPING RATIO FROM THE ROOT-LOCUS FOR CONTINUOUS SYSTEMS

13.39. Prove Equation (*13.18*).

The roots of $s^2 + 2\zeta\omega_n s + \omega_n^2$ are $s_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$. Then

$$|s_1| = |s_2| = \sqrt{\zeta^2\omega_n^2 + \omega_n^2(1-\zeta^2)} = \omega_n$$

and

$$\arg s_{1,2} = \mp \tan^{-1}\left(\sqrt{1-\zeta^2}/\zeta\right) = 180° \pm \theta$$

or $s_{1,2} = \omega_n \angle 180° \pm \theta$. Thus $\cos\theta = \zeta\omega_n/\omega_n = \zeta$.

13.40. Determine the positive value of gain which results in a damping ratio of 0.55 for the complex poles on the root-locus shown in Fig. 13-12.

The angle of the desired poles is $\theta = \cos^{-1} 0.55 = 56.6°$. A line drawn from the origin at an angle of 56.6° with the negative real axis intersects the root-locus of Fig. 13-12 at $K = 7$.

13.41. Find the damping ratio of the complex poles of Problem 13.26 for $K = 3.5$.

A line drawn from the root-locus at $K = 3.5$ to the origin makes an angle of 53° with the negative real axis. Hence the damping ratio of the complex poles is $\zeta = \cos 53° = 0.6$.

Supplementary Problems

13.42. Determine the angle and magnitude of

$$GH = \frac{16(s+1)}{s(s+2)(s+4)}$$

at the following points in the s-plane: (*a*) $s = j2$, (*b*) $s = -2 + j2$, (*c*) $s = -4 + j2$, (*d*) $s = -6$, (*e*) $s = -3$.

13.43. Determine the angle and magnitude of

$$GH = \frac{20(s+10+j10)(s+10-j10)}{(s+10)(s+15)(s+25)}$$

at the following points in the s-plane: (*a*) $s = j10$, (*b*) $s = j20$, (*c*) $s = -10 + j20$, (*d*) $s = -20 + j20$, (*e*) $s = -15 + j5$.

13.44. For each transfer function, find the breakaway points on the root-locus:

(*a*) $GH = \dfrac{K}{s(s+6)(s+8)}$, (*b*) $GH = \dfrac{K(s+5)}{(s+2)(s+4)}$, (*c*) $GH = \dfrac{K(s+1)}{s^2(s+9)}$.

13.45. Find the departure angle of the root-locus from the pole at $s = -10 + j10$ for

$$GH = \frac{K(s+8)}{(s+14)(s+10+j10)(s+10-j10)} \qquad K > 0$$

13.46. Find the departure angle of the root-locus from the pole at $s = -15 + j9$ for

$$GH = \frac{K}{(s+5)(s+10)(s+15+j9)(s+15-j9)} \qquad K > 0$$

13.47. Find the arrival angle of the root-locus to the zero at $s = -7 + j5$ for

$$GH = \frac{K(s+7+j5)(s+7-j5)}{(s+3)(s+5)(s+10)} \qquad K > 0$$

13.48. Construct the root-locus for $K > 0$ for the transfer function of Problem 13.44(a).

13.49. Construct the root-locus for $K > 0$ for the transfer function of Problem 13.44(c).

13.50. Construct the root-locus for $K > 0$ for the transfer function of Problem 13.45.

13.51. Construct the root-locus for $K > 0$ for the transfer function of Problem 13.46.

13.52. Determine the gain and phase margins for the system with the open-loop transfer function of Problem 13.46 if the gain factor K is set equal to 20,000.

Answers to Some Supplementary Problems

13.42. (a) $\arg GH = -99°$, $|GH| = 1.5$; (b) $\arg GH = -153°$, $|GH| = 2.3$; (c) $\arg GH = -232°$, $|GH| = 1.8$; (d) $\arg GH = 0°$, $|GH| = 1.7$; (e) $\arg GH = -180°$, $|GH| = 10.7$

13.43. (a) $\arg GH = -38°$, $|GH| = 0.68$; (b) $\arg GH = -40°$, $|GH| = 0.37$; (c) $\arg GH = -41°$, $|GH| = 0.60$; (d) $\arg GH = -56°$, $|GH| = 0.95$; (e) $\arg GH = +80°$, $|GH| = 6.3$

13.44. (a) $\sigma_b = -2.25, -7.07$; (b) $\sigma_b = -3.27, -6.73$; (c) $\sigma_b = 0, -3$

13.45. $\theta_D = 124°$

13.46. $\theta_D = 193°$

13.47. $\theta_A = 28°$

13.52. Gain margin = 3.7; phase margin = 102°

Chapter 14

Root-Locus Design

14.1 THE DESIGN PROBLEM

The root-locus method can be quite effective in the design of either continuous or discrete-time feedback control systems, because it graphically illustrates the variation of the system closed-loop poles as a function of the open-loop gain factor K. In its simplest form, design is accomplished by choosing a value of K which results in satisfactory closed-loop behavior. This is called *gain factor compensation* (also see Section 12.2). Specifications on allowable steady state errors usually take the form of a minimum value of K, expressed in terms of error constants, for example, K_p, K_v, and K_a (Chapter 9). If it is not possible to meet all system specifications using gain factor compensation alone, other forms of compensation can be added to the system to alter the root-locus as needed, for example, lag, lead, lag-lead networks, or PID controllers.

In order to accomplish system design in the s-plane or the z-plane using root-locus techniques, it is necessary to interpret the system specifications in terms of desired pole-zero configurations.

Digital computer programs for constructing root-loci can be very helpful in system design, as well as analysis as indicated in Chapter 13.

EXAMPLE 14.1. Consider the design of a continuous unity feedback system with the plant $G = K/(s+1)(s+3)$ and the following specifications: (1) Overshoot less than 20%, (2) $K_p \geq 4$, (3) 10 to 90% rise time less than 1 sec.

The root-locus for this system is shown in Fig. 14-1. The system closed-loop transfer function may be written as

$$\frac{C}{R} = \frac{K}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

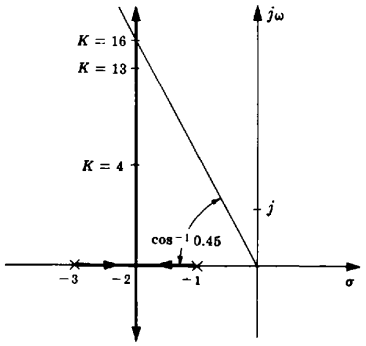


Fig. 14-1

where ζ and ω_n can be determined from the root-locus for a given value of K. In order to satisfy the first specification, ζ must be greater than 0.45 (see Fig. 3-4). Then from the root-locus we see that K must be less than 16 (see Section 13.12). For this system, K_p is given by $K/3$. Thus in order to satisfy the second specification, K must be greater than 12. The rise time is a function of both ζ and ω_n. Suppose a trial value of $K = 13$ is chosen. In this case, $\zeta = 0.5$, $\omega_n = 4$, and the rise time is 0.5 sec. Hence all the specifications can be met by setting $K = 13$. Note that if the specification on K_p was greater than 5.33, or the specification on rise time was less than 0.34 sec, all the specifications could not be met by simply adjusting the open-loop gain factor.

14.2 CANCELLATION COMPENSATION

If the pole-zero configuration of the plant is such that the system specifications cannot be met by an adjustment of the open-loop gain factor, a more complicated cascade compensator, as shown in Fig. 14-2, can be added to the system to cancel some or all of the poles and zeros of the plant. Due to realizability considerations, the compensator must have no more zeros than poles. Consequently, when poles of the plant are cancelled by zeros of the compensator, the compensator also adds new poles to the forward-loop transfer function. The philosophy of this compensation technique is then to replace undesirable with desirable poles.

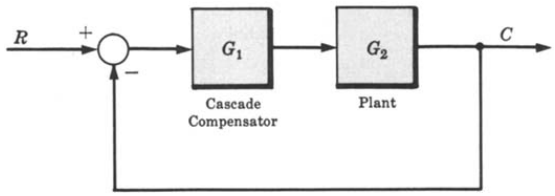


Fig. 14-2

The difficulty encountered in applying this scheme is that it is not always apparent what open-loop pole-zero configuration is desirable from the standpoint of meeting specifications on closed-loop system performance.

Some situations where cancellation compensation can be used to advantage are the following:

1. If the specifications on system rise time or bandwidth cannot be met without compensation, cancellation of low-frequency poles and replacement with high-frequency poles is helpful.
2. If the specifications on allowable steady state errors cannot be met, a low-frequency pole can be cancelled and replaced with a lower-frequency pole, yielding a larger forward-loop gain at low frequencies.
3. If poles with small damping ratios are present in the plant transfer function, they may be cancelled and replaced with poles which have larger damping ratios.

14.3 PHASE COMPENSATION: LEAD AND LAG NETWORKS

A cascade compensator can be added to a system to alter the phase characteristics of the open-loop transfer function in a manner which favorably affects system performance. These effects were illustrated in the frequency domain for lead, lag, and lag-lead networks using Polar Plots in Chapter 12, Sections 12.4 through 12.7, which summarize the general effects of these networks.

The pole-zero maps of continuous system lead and lag networks are shown in Figs. 14-3 and 14-4. Note that a lead network makes a positive, and a lag network a negative phase contribution. A lag-lead

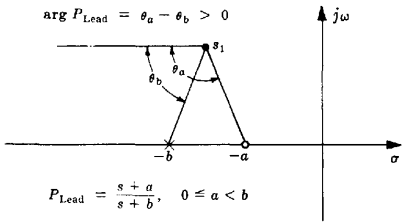


Fig. 14-3

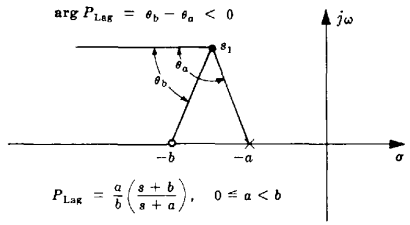


Fig. 14-4

network may be obtained by appropriately combining a lag and a lead network in series, or from the implementation described in Problem 6.14.

Since the compensated system root-locus is determined by the points in the complex plane for which the phase angle of $G = G_1G_2$ is equal to $-180°$, the branches of the locus can be moved by proper selection of the phase angle contributed by the compensator. In general, lead compensation has the effect of moving the loci to the left.

EXAMPLE 14.2. The phase lead compensator $G_1 = (s + 2)/(s + 8)$ alters the root-locus of the system with the plant $G_2 = K/(s + 1)^2$, as illustrated in Fig. 14-5.

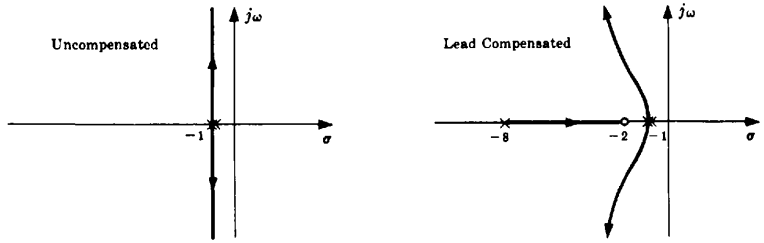


Fig. 14-5

EXAMPLE 14.3. The use of *simple lag compensation* (one pole at -1, no zero) to alter the breakaway angles of a root-locus from a pair of complex poles is illustrated in Fig. 14-6.

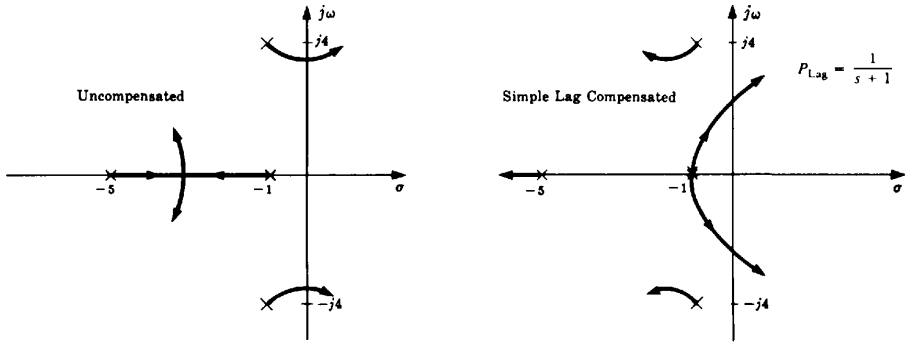


Fig. 14-6

14.4 MAGNITUDE COMPENSATION AND COMBINATIONS OF COMPENSATORS

Compensation networks may be employed to alter the closed-loop magnitude characteristic ($|(C/R)(\omega)|$) of a feedback control system. The low-frequency characteristic can be modified by addition of a low-frequency pole-zero pair, or **dipole**, in such a manner that high-frequency behavior is essentially unaltered.

EXAMPLE 14.4. The continuous system root-locus for $GH = K/s(s + 2)^2$ is shown in Fig. 14-7.

Let us assume that this system has a satisfactory transient response with $K = 3$, but the resulting velocity error constant, $K_v = 0.75$, is too small. We can increase K_v to 5 without seriously affecting the transient response by adding the compensator $G_1 = (s + 0.1)/(s + 0.015)$ since

$$K_v' = K_v G_1(0) = \frac{0.75(0.1)}{0.015} = 5$$

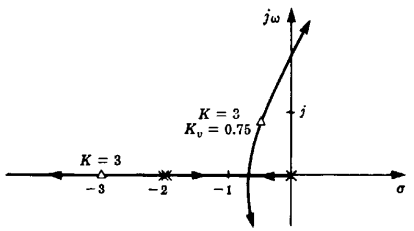


Fig. 14-7

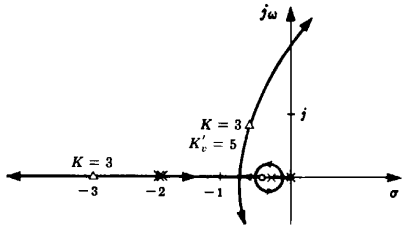


Fig. 14-8

The resulting root-locus is shown in Fig. 14-8. The high-frequency portion of the root-locus and the transient response are essentially unaffected because the closed-loop transfer function has a low-frequency pole-zero pair which approximately cancel each other.

A low-frequency dipole for magnitude compensation of continuous systems can be synthesized with the pole at the origin using a proportional plus integral (PI) compensator, as shown in Fig. 14-9, with transfer function

$$G_1 = \frac{s + K_I}{s}$$

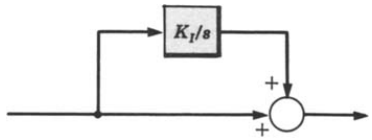


Fig. 14-9

Combinations of various compensation schemes are sometimes needed to satisfy competing requirements on steady state and transient response performance specifications, as illustrated in the following example. This example, solved by root-locus methods, is a rework of a design problem solved by Nyquist methods in Example 12.7, and Bode methods in Example 16.6.

EXAMPLE 14.5. Our goal is to determine an appropriate compensator $G_1(z)$ for the discrete-time unity feedback system with

$$G_2(z) = \frac{3(z+1)(z+\frac{1}{3})}{8z(z+0.5)}$$

The resulting closed-loop system must satisfy the following performance specifications:

1. Steady state error less than or equal to 0.02 for a unit ramp input
2. Phase margin $= \phi_{PM} \geq 30°$
3. Gain crossover frequency $\omega_1 \geq 10$ rad/sec, where $T = 0.1$ sec.

In order to have a finite steady state error with a ramp input, the system must be type 1. The compensation must therefore provide a pole at $z = 1$. Consider the compensator

$$G_1' = \frac{K_1}{z-1}$$

The forward-loop transfer function then becomes

$$G_1'G_2(z) = \frac{3K_1(z+1)(z+\frac{1}{3})}{8(z-1)z(z+0.5)}$$

From Section 9.9, the velocity error coefficient is

$$K_v = \frac{3K_1(1+1)(1+\frac{1}{3})}{8(1)(1+0.5)} = 0.667K_1$$

Now, in order for the system to have a steady state error of less than 0.02 with a ramp input, we must have $K_v \geq 5$, or $K_1 \geq 7.5$. To investigate the effects of added gain, we consider the root-locus for

$$G_1'G_2(z) = \frac{K(z+1)(z+\frac{1}{3})}{z(z+0.5)(z-1)}$$

where $K = 3K_1/8$. This root-locus was constructed in Problem 13.31 and is repeated in Fig. 14-10.

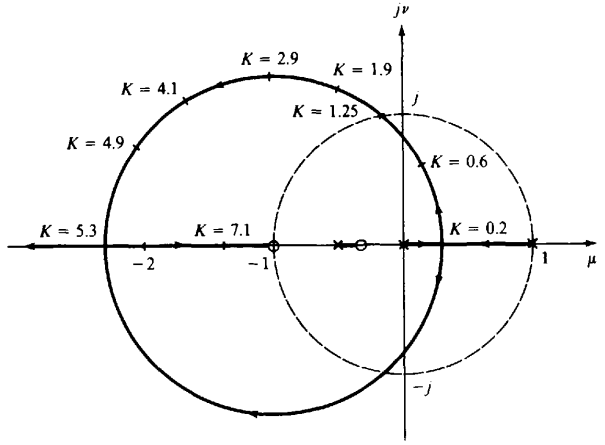


Fig. 14-10

At the point $z = -0.18 + j0.98$ where the root-locus crosses the unit circle, $\omega_n T = 1.75$ rad and $K = 1.25$ ($K_1 = 8K/3 = 3.33$). Since this is less than the gain $K_1 = 7.5$ needed to make $K_v = 5$, simple gain factor compensation is insufficient.

The next step is to evaluate the magnitude and phase of $G_1'G_2(z)$ at the required minimum gain crossover frequency, $\omega_1 = 10$, or $\omega_1 T = 1$ rad. This is the point $z = e^{j\omega T} = e^j$ on the unit circle. At this point, $|G_1'G_2(e^j)| = 1.66K$ and $\arg G_1'G_2(e^j) = -142.5°$. If the gain K were adjusted so that $|G_1'G_2(e^j)| = 1$, that is, $K = 0.6$, the phase margin would be $(180 - 142.5)° = 37.5°$ and the 30° requirement would be met. This requires that $K_1 = 8K/3 = 1.6$, and the velocity constant becomes $K_v = 0.667K_1 = 1.067$.

To complete the design, additional gain must be added to increase the velocity constant to the required value of 5 at low frequencies, without significantly altering the desired high-frequency characteristics obtained so far. This requires an additional gain of $5/1.067 = 4.69$, which can be supplied by a lag compensator. The lag compensator should have a gain at $z = 1$ that is 4.69 times as large as the gain at $\omega T = 1$, without adding more than 7.5° phase lag at $\omega T = 1$, to satisfy the requirement for $\phi_{PM} \geq 30°$. If a value of 0.97 is chosen for the pole of the lag compensator, the zero should be located so that

$$P_{Lag} = \frac{1 - z_1}{1 - 0.97} \geq 4.69$$

or, $z_1 \leq 0.86$. If we set $z_1 = 0.86$, then

$$|P_{Lag}| = \left|\frac{z - 0.86}{z - 0.97}\right| = \begin{cases} 4.7 & \text{for } z = 1 \\ 0.95 & \text{for } z = e^j \end{cases} \qquad (\omega T = 1)$$

and

$$\arg P_{Lag} = \arg\left(\frac{e^j - 0.86}{e^j - 0.97}\right) = -6.25° \quad \text{for } z = e^j$$

The compensator then becomes

$$G_1 = \frac{K_1(z - 0.86)}{(z - 0.97)(z - 1)}$$

Finally, for $\omega_1 T = 1$, we need $|G_1G_2(e^j)| = 1$, so $K_1 = 1.60/0.95 = 1.68$, to account for the gain of the lag

compensator at $\omega T = 1$. The completed compensator is

$$G_1 = \frac{1.68(z - 0.86)}{(z - 0.97)(z - 1)}$$

which is nearly the same design obtained by Nyquist methods in Example 12.7.

14.5 DOMINANT POLE-ZERO APPROXIMATIONS

The root-locus method offers the advantage of a graphical display of the system closed-loop poles and zeros. Theoretically, the designer can determine the system response characteristics from the closed-loop pole-zero map. Practically, however, this task becomes increasingly difficult for systems with four or more poles and zeros. In some cases the problem can be considerably simplified if the response is dominated by two or three poles and zeros.

Effects on System Time Responses

The influence of a particular pole (or pair of complex poles) on the response is mainly determined by two factors: the relative rate of decay of the transient term due to that pole, and the relative magnitude of the residue at the pole.

For **continuous systems**, the real part σ of the pole p determines the rate at which the transient term due to that pole decays; the larger σ, the faster the rate of decay. The relative magnitude of the residue determines the percentage of the total response due to that particular pole.

EXAMPLE 14.6. Consider a system with closed-loop transfer function

$$\frac{C}{R} = \frac{5}{(s + 1)(s + 5)}$$

The step response of this system is

$$c(t) = 1 - 1.25e^{-t} + 0.25e^{-5t}$$

The term in the response due to the pole at $s_1 = \sigma_1 = -5$ decays five times as fast as the term due to the pole at $s_2 = \sigma_2 = -1$. Furthermore, the residue at the pole at $s_1 = -5$ is only $\frac{1}{5}$ that of the one at $s_2 = -1$. Therefore for most practical purposes the effect of the pole at $s_1 = -5$ can be ignored and the system approximated by

$$\frac{C}{R} \cong \frac{1}{s + 1}$$

The pole at $s_1 = -5$ has been removed from the transfer function and the numerator has been adjusted to maintain the same steady state gain ($(C/R)(0) = 1$). The response of the approximate system is $c(t) = 1 - e^{-t}$.

EXAMPLE 14.7. The system with the closed-loop transfer function

$$\frac{C}{R} = \frac{5.5(s + 0.91)}{(s + 1)(s + 5)}$$

has the step response

$$c(t) = 1 + 0.125e^{-t} - 1.125e^{-5t}$$

In this case, the presence of a zero close to the pole at -1 significantly reduces the magnitude of the residue at that pole. Consequently, it is the pole at -5 which now dominates the response of the system. The closed-loop pole and zero effectively cancel each other and $(C/R)(0) = 1$ so that an approximate transfer function is

$$\frac{C}{R} \cong \frac{5}{s + 5}$$

and the corresponding approximate step response is $c \cong 1 - e^{-5t}$.

For **discrete-time systems** with distinct (nonrepeated) poles $p_1, p_2, \ldots$, the transient portion $y_T(k)$ of the response due to a pole p has the form $y_T(k) = p^k$, $k = 0, 1, 2, \ldots$ (see Table 4.2). Therefore each

successive time sample is equal to the previous sample multiplied by p, that is,

$$y_T(k + 1) = py_T(k)$$

The magnitude of a distinct pole therefore determines the decay rate of the transient response, with the decay rate inversely proportional to $|p|$: the smaller the magnitude, the faster the rate of decay. For example, poles near the unit circle decay more slowly than poles near the origin, since their magnitudes are smaller.

For systems with repeated poles, the analysis is more complicated and approximations may not be appropriate.

EXAMPLE 14.8. The discrete system with closed-loop transfer function

$$\frac{C}{R} = \frac{0.45z}{(z - 0.1)(z - 0.5)}$$

has the step response

$$c(k) = 1 - 1.125(0.5)^k + 0.125(0.1)^k \qquad k = 0, 1, 2, \ldots$$

For the term in the response due to the pole at $z = 0.1$, the sample value at time k is only 10% of the sample value at time $k - 1$, and it therefore decays five times faster than the term due to the pole at $z = 0.5$. The magnitude of the residue at $z = 0.1$ is 0.125, which is one-ninth as large as the magnitude of the residue 1.125 at $z = 0.5$. Consequently, for many practical purposes, the pole at $z = 0.1$ can often be ignored and the system approximated by

$$\frac{C}{R} \cong \frac{0.5}{z - 0.5}$$

where the numerator has been adjusted to maintain the same steady state gain

$$\frac{C}{R}(1) = 1$$

and the zero at $z = 0$ was deleted to maintain one more pole than zeros in the approximate system. This is necessary to give the same initial delay (one sample time) in the approximate system as in the original system. The step response of the approximate system is $c(k) = 1 - (0.5)^k$, $k = 0, 1, 2, \ldots$.

Effects on Other System Characteristics

The effect of a closed-loop real axis *pole* at $-p_r < 0$ on the overshoot and rise time T_r of a continuous system also having complex poles $-p_c, -p_c^*$ is illustrated in Figs. 14-11 and 14-12. For

$$\frac{p_r}{\zeta\omega_n} > 5 \qquad (14.1)$$

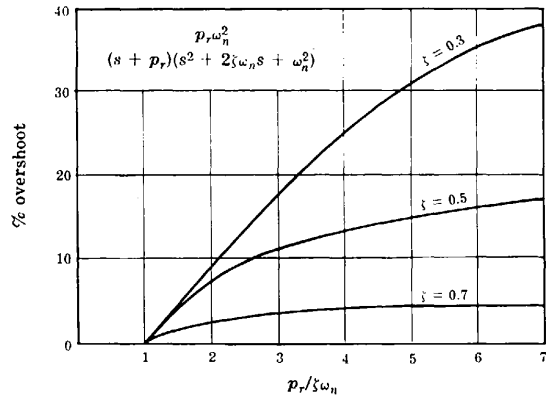


Fig. 14-11

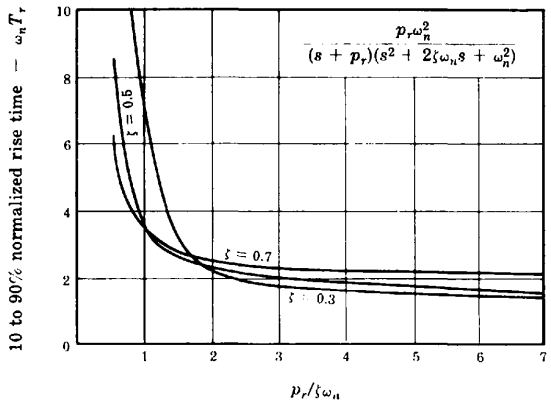


Fig. 14-12

the overshoot and rise time approach that of a second-order system containing only complex poles (see Fig. 3-4). Therefore p_r can be neglected in determining overshoot and rise time if $\zeta > 0.5$ and

$$p_r > 5|\text{Re } p_c| = 5\zeta\omega_n \tag{14.2}$$

There is no overshoot if

$$p_r \leq |\text{Re } p_c| = \zeta\omega_n \tag{14.3}$$

and the rise time approaches that of a first-order system containing only the real axis pole.

The effect of a closed-loop real axis *zero* at $-z_r < 0$ on the overshoot and rise time T_r of a continuous system also having complex poles $-p_c$, $-p_c^*$ is illustrated in Figs. 14-13 and 14-14. These graphs show that z_r can be neglected in determining overshoot and rise time if $\zeta > 0.5$ and

$$z_r > 5|\text{Re } p_c| = 5\zeta\omega_n \tag{14.4}$$

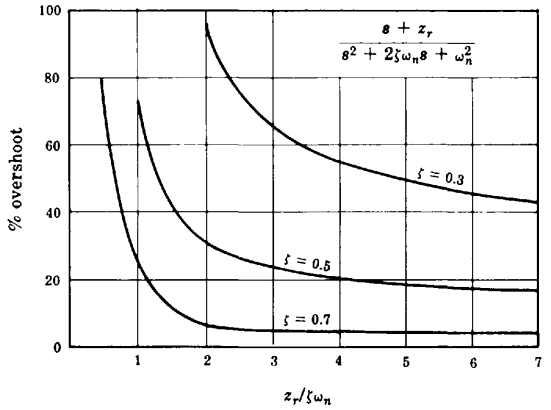


Fig. 14-13

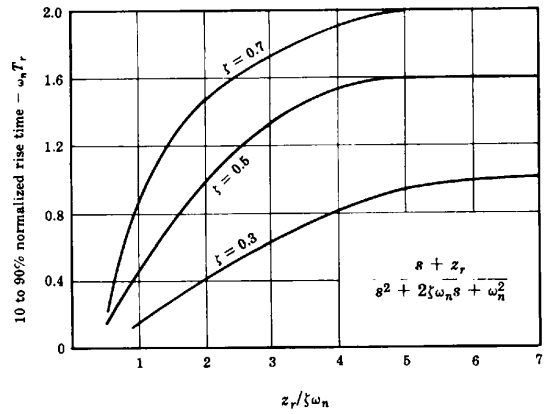


Fig. 14-14

EXAMPLE 14.9. The closed-loop transfer function of a particular continuous system is represented by the pole-zero map shown in Fig. 14-15. Given that the steady state gain $(C/R)(j0) = 1$, a dominant pole-zero approximation is

$$\frac{C}{R} \cong \frac{4}{s^2 + 2s + 4}$$

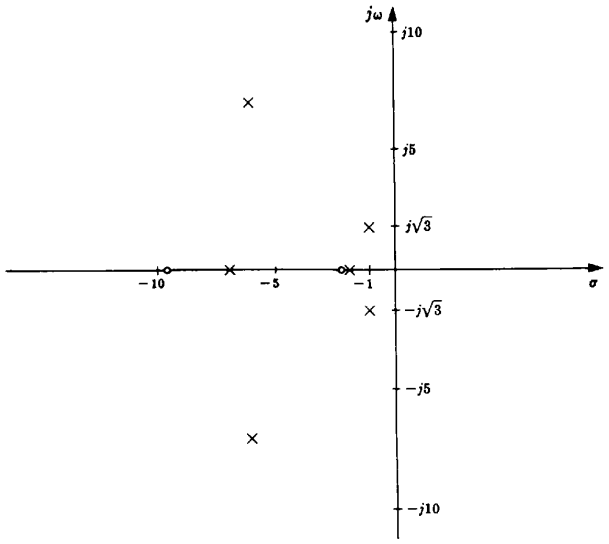


Fig. 14-15

This is a reasonable approximation because the pole and zero near $s = -2$ effectively cancel each other and all other poles and zeros satisfy Equations (*14.2*) and (*14.4*) with $-p_c = -1 + j\sqrt{3}$ and $\zeta = 0.5$.

14.6 POINT DESIGN

If a desired closed-loop pole position p_1 can be determined from the system specifications, the system root-locus may be altered to ensure that a branch of the locus will pass through the required point p_1. The specification of a closed-loop pole at a particular point in the complex plane is called **point design**. The technique is carried out using phase and magnitude compensation.

EXAMPLE 14.10. Consider the continuous plant

$$G_2 = \frac{K}{s(s+2)^2}$$

The closed-loop response must have a 10 to 90% rise time less than 1 sec, and an overshoot less than 20%. We observe from Fig. 3-4 that these specifications are met if the closed-loop system has a dominant two-pole configuration with $\zeta = 0.5$ and $\omega_n = 2$. Thus p_1 is chosen at $-1 + j\sqrt{3}$, which is a solution of

$$p_1^2 + 2\zeta\omega_n p_1 + \omega_n^2 = 0$$

for $\zeta = 0.5$ and $\omega_n = 2$. Clearly, $p_1^* = -1 - j\sqrt{3}$ is the remaining solution of this quadratic equation. The orientation of p_1 with respect to the poles of G_2 is shown in Fig. 14-16.

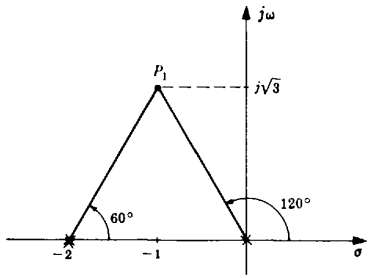


Fig. 14-16

The phase angle of G_2 is $-240°$ at p_1. In order for a branch of the root-locus to pass through p_1, the system must be modified so that the phase angle of the compensated system is $-180°$ at p_1. This can be accomplished by adding a cascade lead network having a phase angle of $240° - 180° = 60°$ at p_1, which is satisfied by

$$G_1 = P_{\text{Lead}} = \frac{s+1}{s+4}$$

as shown in the pole-zero map of the compensated open-loop transfer function G_1G_2 in Fig. 14-17. The closed-loop pole can now be located at p_1 by choosing a value for K which satisfies the root-locus magnitude criterion. Solution of Equation (*13.5*) yields $K = 16$. The root-locus or closed-loop pole-zero map of the compensated system should be sketched to check the validity of the dominant two-pole assumption. Figure 14-18 illustrates that the poles at p_1 and p_1^* dominate the response.

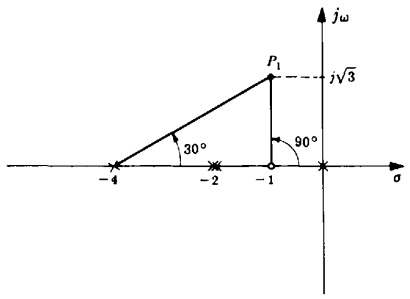


Fig. 14-17

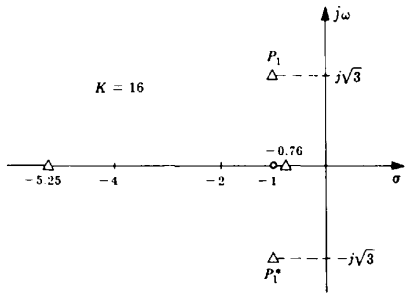


Fig. 14-18

14.7 FEEDBACK COMPENSATION

Addition of compensation elements to a feedback path of a control system can be employed in root-locus design in a manner similar to that discussed in the preceding sections. The compensation elements affect the root-locus of the open-loop transfer function in the same manner. But, although the root-locus is the same when the compensator is in either the forward or feedback path, the closed-loop transfer function may be significantly different. It was shown in Problem 13.34 that feedback *zeros* do not appear in the closed-loop transfer function, while feedback *poles* become zeros of the closed-loop transfer function (assuming no cancellations).

EXAMPLE 14.11. Suppose a feedback compensator were added to a continuous system with the forward transfer function

$$G = \frac{K}{(s+1)(s+4)(s+5)}$$

in an attempt to cancel the pole at -1 and replace it with a pole at -6. Then the compensator would be $H = (s+1)/(s+6)$, GH would be given by $GH = K/(s+4)(s+5)(s+6)$ and the closed-loop transfer function would become

$$\frac{C}{R} = \frac{K(s+6)}{(s+1)[(s+4)(s+5)(s+6)+K]}$$

Although the pole at -1 is cancelled from GH, it reappears as a *closed-loop* pole. Furthermore, the feedback pole at -6 becomes a closed-loop zero. Consequently, *the cancellation technique does not work with a compensator in the feedback path.*

EXAMPLE 14.12. The continuous system block diagram in Fig. 14-19 contains two feedback paths.

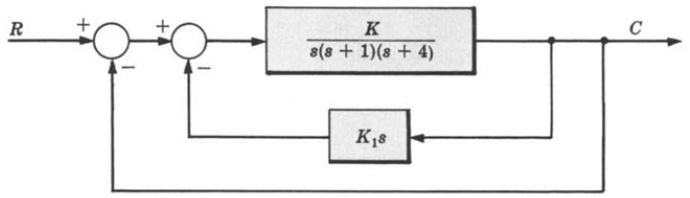


Fig. 14-19

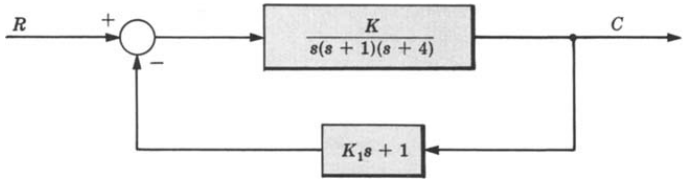


Fig. 14-20

These two paths may be combined, as shown in Fig. 14-20.

In this representation the feedback path contains a zero at $s = -1/K_1$. This zero appears in GH and consequently affects the root-locus. However, it does not appear in the closed-loop transfer function, which contains three poles no matter where the zero is located.

The fact that feedback zeros do not appear in the closed-loop transfer function may be used to advantage in the following manner. If closed-loop poles are desired at certain locations in the complex plane, feedback zeros can be placed at these points. Since branches of the root-locus will terminate on these zeros, the desired closed-loop pole locations can be obtained by setting the open-loop gain factor sufficiently high.

EXAMPLE 14.13. The continuous system feedback compensator

$$H = \frac{s^2 + 2s + 4}{(s+6)^2}$$

is added to the system with the forward-loop transfer function

$$G = \frac{K}{s(s+2)}$$

in order to guarantee that the dominant closed-loop poles will be near $s = -1 \pm j\sqrt{3}$. The resulting root-locus is shown in Fig. 14-21.

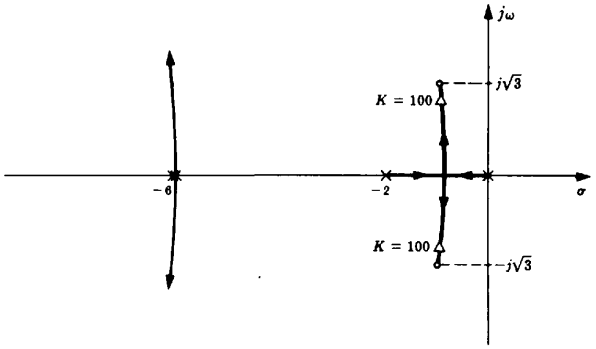


Fig. 14-21

If K is set at 100, the closed-loop transfer function is

$$\frac{C}{R} = \frac{100(s+6)^2}{(s^2 + 1.72s + 2.96)(s^2 + 12.3s + 135)}$$

and the dominant complex pole pair $s_{1,2} = 0.86 \pm j1.5$ are sufficiently close to $-1 \pm j\sqrt{3}$.

Solved Problems

GAIN FACTOR COMPENSATION

14.1. Determine the value of the gain factor K for which the system with the open-loop transfer function

$$GH = \frac{K}{s(s+2)(s+4)}$$

has closed-loop poles with a damping ratio $\zeta = 0.5$.

The closed-loop poles will have a damping ratio of 0.5 when they make an angle of 60° with the negative real axis [Equation (*13.18*)]. The desired value of K is determined at the point where the root-locus crosses the $\zeta = 0.5$ line in the s-plane. A sketch of the root-locus is shown in Fig. 14-22. The desired value of K is 8.3.

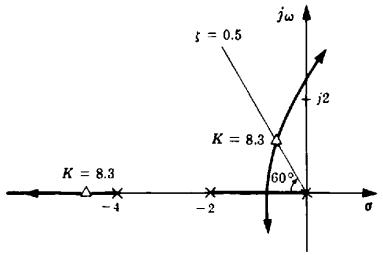


Fig. 14-22

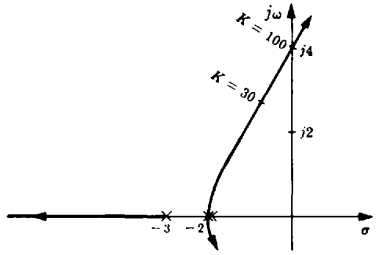


Fig. 14-23

14.2. Determine a value of K for which the system with the open-loop transfer function

$$GH = \frac{K}{(s+2)^2(s+3)}$$

satisfies the following specifications: (*a*) $K_p \geq 2$, (*b*) gain margin ≥ 3.

For this system, K_p is equal to $K/12$. Hence, in order to satisfy the first specification, K must be greater than 24. The value of K at the $j\omega$-axis crossover of the root-locus is equal to 100, as shown in Fig. 14-23. Then, in order to satisfy the second specification, K must be less than $100/3 = 33.3$. A value of K that will satisfy both specifications is 30.

14.3. Determine a gain factor K for which the system in Example 13.11 has a gain margin of 2.

As shown in Fig. 13-15, the gain at the stability boundary is $K = 1$. Therefore, in order to have a gain margin of 2, K must be 0.5.

CANCELLATION COMPENSATION

14.4. Can right-half s-plane poles of a plant be effectively cancelled by a compensator with a right-half s-plane zero?

No. For example, suppose a particular plant has the transfer function

$$G_2 = \frac{K}{s-1} \qquad K > 0$$

and a cascade compensator is added with the transfer function $G_1 = (s - 1 + \epsilon)/(s+1)$. The ϵ term in the transfer function represents any small error between the desired zero location at $+1$ and the actual location. The closed-loop transfer function is then

$$\frac{C}{R} = \frac{K(s - 1 + \epsilon)}{s^2 + Ks + K\epsilon - K - 1}$$

By applying the Hurwitz or Routh Stability Criterion (Chapter 5) to the denominator of this transfer function, it can be seen that the system is unstable for any value of K if ϵ is less than $(1 + K)/K$, which is usually the case because ϵ represents the error in the desired zero location.

14.5. For the discrete-time unity feedback system with forward-loop transfer function

$$G_2 = \frac{z+1}{z(z-1)}$$

determine a compensator G_1 that provides a deadbeat response for the closed-loop system.

For a deadbeat response (Section 10.8), we want all closed-loop poles at $z = 0$. A pole-zero map of the system is shown in Fig. 14-24(a). If we cancel the pole at $z = 0$ and the zero at $z = -1$, the root-locus will go through $z = 0$, as shown in Fig. 14-24(b). The resulting compensator is then

$$G_1 = \frac{z}{z+1}$$

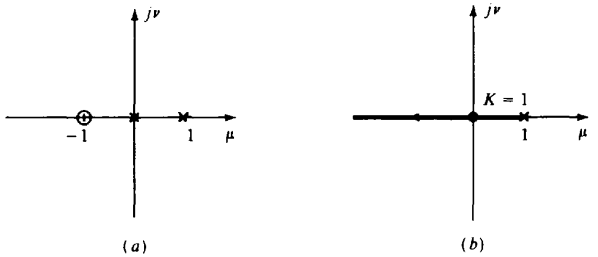


Fig. 14-24

and the closed-loop transfer function is

$$\frac{C}{R} = \frac{G_1G_2}{1+G_1G_2} = \frac{1}{z}$$

PHASE COMPENSATION

14.6. It is desired to add to a system a compensator with a *zero* at $s = -1$ to produce 60° phase lead at $s = -2 + j3$. How can the proper location of the *pole* be determined?

With reference to Fig. 14-3, we want the phase contribution of the network to be $\theta_a - \theta_b = 60°$. From Fig. 14-25, $\theta_a = 108°$. Hence $\theta_b = \theta_a - 60° = 48°$ and the pole should be located at $s = -4.7$, as shown in Fig. 14-25.

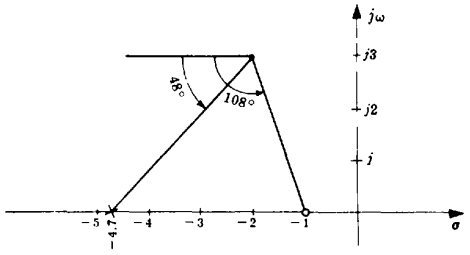


Fig. 14-25

14.7. Determine a compensator that will change the departure angle of the root-locus from the pole at $s = -0.5 + j$ to $-135°$ for the plant transfer function

$$G_2 = \frac{K}{s(s^2+s+1.25)}$$

The departure angle of the uncompensated system is $-27°$. To change this to $-135°$, a lag compensator with 108° phase lag at $s = -0.5 + j$ can be employed. The required amount of phase lag could be supplied by a *simple* lag compensator (one pole, no zero) with a pole at $s = -0.18$, as shown in Fig. 14-26(a), or by two simple lags in cascade with two poles at $s = -1.22$, as shown in Fig. 14-26(b).

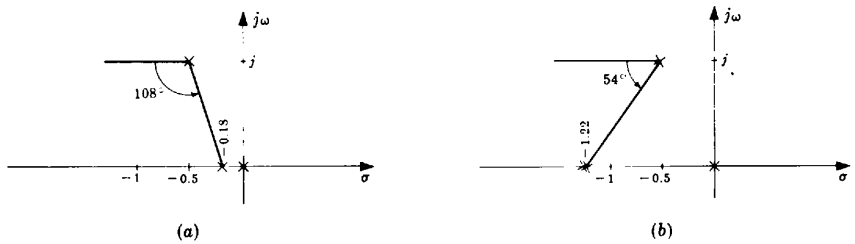


Fig. 14-26

14.8. Determine a compensator for the discrete-time system with

$$GH(z) = \frac{K}{z(z-1)}$$

that provides a phase crossover frequency ω_π such that $\omega_\pi T = \pi/2$ rad.

Arg GH at $z = e^{j\pi/2} = j$ is determined from the pole-zero map in Fig. 14-27 as $-225°$. In order for the root-locus to pass through this point, we need to add 45° of phase lead, so that arg $GH = \pm 180°$. This can be provided by the compensator

$$P_{\text{Lead}}(z) = \frac{z}{z+1}$$

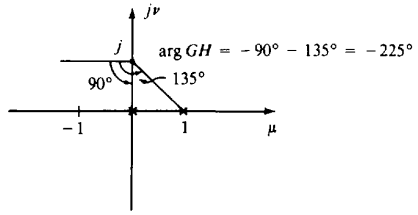


Fig. 14-27

The zero at $z = 0$ provides 90° of phase lead and the pole at $z = -1$ provides 45° of lag, resulting in a net lead of 45°.

MAGNITUDE COMPENSATION

14.9. In Example 14.4, the velocity error constant K_v was increased by a factor of $6\frac{2}{3}$ without increasing the gain factor. How was this accomplished?

It was assumed that the compensator G_1 had a high-frequency gain of 1 and a low-frequency (d.c.) gain of $6\frac{2}{3}$. This compensator cannot be mechanized passively because a passive lag compensator has a d.c. gain of 1. Consequently, G_1 must include an amplifier. An alternative method would be to let G_1 be the passive lag compensator

$$G_1' = \frac{0.015}{0.1}\left(\frac{s+0.1}{s+0.015}\right)$$

and then amplify the gain factor by $6\frac{2}{3}$. However, when root-locus techniques are employed it is usually more convenient to assume the compensator just adds a pole and zero, as was done in Example 14.4. Appropriate adjustments can be made in the final stages of design to achieve the simplest and/or least expensive compensator mechanization.

DOMINANT POLE-ZERO APPROXIMATIONS

14.10. Determine the overshoot and rise time of the system with the transfer function

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

For this system, $\omega_n = 1$, $\zeta = 0.5$, $p_r = 1$, and $p_r/\zeta\omega_n = 2$. From Fig. 14-11 the percentage overshoot is about 8%. The rise time from Fig. 14-12 is 2.4 sec. The corresponding numbers for a system with the complex poles only are 18% and 1.6 sec. Thus the real axis pole reduces the overshoot and slows down the response.

14.11. Determine the overshoot and rise time of the system with the transfer function

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

For this system $\omega_n = 1$, $\zeta = 0.5$, $z_r = 1$, and $z_r/\zeta\omega_n = 2$. From Fig. 14-13 the percentage overshoot is 31%. From Fig. 14-14 the 10 to 90% rise time is 1.0 sec. The corresponding numbers for a system without the zero are 18% and 1.6 sec. The real axis zero thus increases the overshoot and decreases the rise time, that is, speeds up the response.

14.12. What is a suitable dominant pole-zero approximation for the following system?

$$\frac{C}{R} = \frac{2(s+8)}{(s+1)(s^2+2s+3)(s+6)}$$

The real axis pole at $s = -6$ and the real axis zero at $s = -8$ satisfy Equations (*14.2*) and (*14.4*), respectively, with regard to the complex poles ($\zeta\omega_n = 1$ and $\zeta > 0.5$) and therefore may be neglected. The real axis pole at $s = -1$ and the complex poles cannot be neglected. Hence a suitable approximation (with the same d.c. gain) is

$$\frac{C}{R} = \frac{8}{3(s+1)(s^2+2s+3)}$$

14.13. Determine a dominant pole approximation for the discrete-time system with transfer function

$$\frac{C}{R} = \frac{0.16}{(z-0.2)(z-0.8)}$$

The step response is given by

$$c(k) = 1 - 1.33(0.8)^k + 0.33(0.2)^k \qquad k = 0,1,2,\ldots$$

The magnitude 0.33 of the residue at $z = 0.2$ is four times smaller than the magnitude 1.33 of the residue at $z = 0.8$. Also, the transient response due to the pole at $z = 0.2$ decays $0.8/0.2 = 4$ times faster than that for the pole at $z = 0.8$. Thus the approximate closed-loop system should only have a pole at $z = 0.8$. However, to maintain a system response delay of two samples (the original system has two more poles than zeros), it is necessary to add a pole at $z = 0$ to the approximation. Then

$$\frac{C}{R} \cong \frac{0.2}{z(z-0.8)}$$

The step response of the approximate system is

$$c(k) = \begin{cases} 0 & \text{for } k = 0 \\ 1 - 1.25(0.8)^k & \text{for } k > 0 \end{cases}$$

Note that the only effect of the pole at $z = 0$ on the response is to delay it by one sample.

POINT DESIGN

14.14. Determine K, a, and b so that the system with open-loop transfer function

$$GH = \frac{K(s+a)}{(s+b)(s+2)^2(s+4)}$$

has a closed-loop pole at $p_1 = -2 + j3$.

The angle contributed to $\arg GH(s_1)$ by the poles at $s = -2$ and $s = -4$ is $-237°$. To satisfy the angle criterion, the angle contributions of the zero at $s = -a$ and the pole $s = -b$ must total $-180° - (-237°) = 57°$. Since this is a positive angle, the zero must be farther to the right than the pole ($b > a$). Either a or b may be chosen arbitrarily as long as the remaining one can be fixed in the finite left-half s-plane to give a total contribution of 57°. Let a be set equal to 2, resulting in a 90° phase contribution. Then b must be placed where the contribution of the pole is $-33°$. A line drawn from p_1 at 33° intercepts the real axis at $6.6 = b$, as shown in Fig. 14-28.

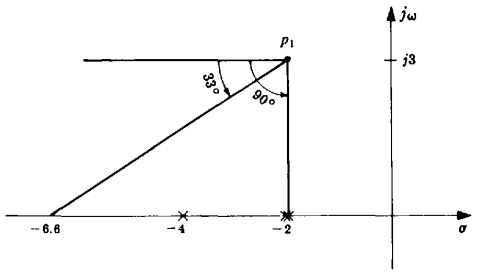


Fig. 14-28

The necessary value of K required to satisfy the magnitude criterion at p_1 can now be computed using the chosen values of a and b. From the following calculation, the required value of K is

$$\left| \frac{(p_1+6.6)(p_1+2)^2(p_1+4)}{(p_1+2)} \right|_{p_1 = -2+j3} = 60$$

14.15. Determine the required compensation for a system with the plant transfer function

$$G_2 = \frac{K}{(s+8)(s+14)(s+20)}$$

to satisfy the following specifications: (a) overshoot $\leq 5\%$, (b) 10 to 90% rise time $T_r \leq 150$ msec, (c) $K_p > 6$.

The first specification may be satisfied with a closed-loop transfer function whose response is dominated by two complex poles with $\zeta \geq 0.7$, as seen from Fig. 3-4. A wide variety of dominant pole-zero configurations can satisfy the overshoot specification; but the two-pole configuration is usually the simplest obtainable form. We also see from Fig. 3-4 that, if $\zeta = 0.7$, the normalized 10 to 90% rise time is about $\omega_n T_r = 2.2$. Thus, in order to satisfy the second specification with $\zeta = 0.7$, we have $T_r = 2.2/\omega_n \leq 0.15$ sec or $\omega_n \geq 14.7$ rad/sec.

But let us choose $\omega_n = 17$ so as to achieve some margin with respect to the rise time specification. Other closed-loop poles, which may appear in the final design, may slow down the response. Thus, in order to satisfy the first two specifications, we shall design the system to have a dominant two-pole response with $\zeta = 0.7$ and $\omega_n = 17$. An s-plane evaluation of $\arg G_2(p_1)$, where $p_1 = -12 + j12$ (corresponding to $\zeta = 0.7$, $\omega_n = 17$), yields $\arg G_2(p_1) = -245°$. Then, to satisfy the angle criterion at p_1, we must compensate the system with phase lead so that the total angle becomes $-180°$. Hence we add a cascade lead compensator with $245° - 180° = 65°$ phase lead at p_1. Arbitrarily placing the zero of the lead compensator at $s = -8$

results in $\theta_a = 108°$ (see Fig. 14-3). Then, since we want $\theta_a - \theta_b = 65°$, $\theta_b = 108° - 65° = 43°$. Drawing a line from p_1 to the real axis at the required θ_b determines the pole location at $s = -25$. Addition of the lead compensator with $a = 8$ and $b = 25$ yields an open-loop transfer function

$$G_2 G_{\text{Lead}} = \frac{K}{(s+14)(s+20)(s+25)}$$

The value of K necessary to satisfy the magnitude criterion at p_1 is $K = 3100$. The resulting positional error constant for this design is $K_p = 3100/(14)(20)(25) = 0.444$, which is substantially less than the specified value of 6 or more. K_p could be increased slightly by trying other design points (higher ω_n); but the required K_p cannot be achieved without some form of low-frequency magnitude compensation. The required increase is $6/0.444 = 13.5$ and may be obtained with a low-frequency lag compensator with $b/a = 13.5$. The only other requirement is that a and b for the lag compensator must be small enough so as not to affect the high-frequency design accomplished with the lead network. That is,

$$\arg P_{\text{Lag}}(p_1) \cong 0$$

Let $b = 1$ and $a = 0.074$. Then the required compensator is

$$G_{\text{Lag}} = \frac{s+1}{s+0.074}$$

To synthesize this compensator using a conventional lag network with the transfer function

$$P_{\text{Lag}} = \frac{0.074(s+1)}{s+0.074}$$

an additional amplifier with a gain of 13.5 is required; equivalently, the design value of K chosen above may be increased by 13.5. With either practical mechanization, the total open-loop transfer function is

$$GH = \frac{3100(s+1)}{(s+0.075)(s+14)(s+20)(s+25)}$$

The closed-loop poles and zeros are shown in Fig. 14-29. The low-frequency pole and zero effectively cancel each other. The real axis pole at $s = -35$ will slightly affect the response of the system because $p_r/\zeta\omega_n$ for this pole is only about 3 [Equation (*14.2*)]. However, reference to Figs. 14-11 and 14-12 verify that the overshoot and rise time are still well within the specifications. If the system had been designed to barely meet the required rise time specification with the dominant two-pole approximation, the presence of the additional pole in the closed-loop transfer function may have slowed the response enough to dissatisfy the specification.

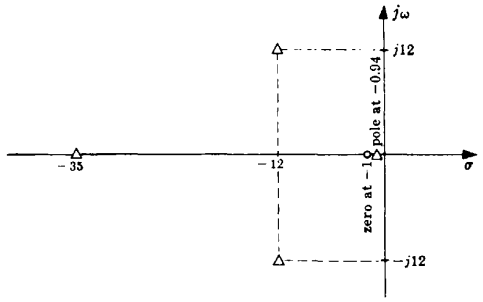


Fig. 14-29

FEEDBACK COMPENSATION

14.16. A positional control system with a tachometer feedback path has the block diagram shown in Fig. 14-30. Determine values of K_1 and K_2 which result in a system design which yields a 10 to 90% rise time of less than 1 sec and an overshoot of less than 20%.

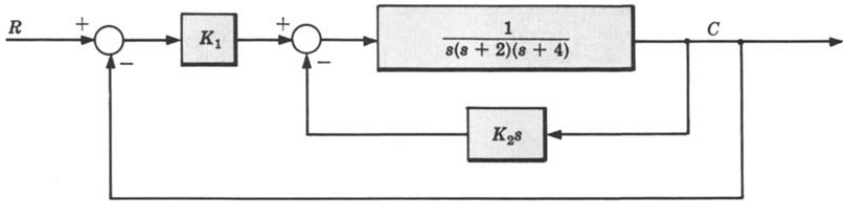


Fig. 14-30

A straightforward way to accomplish this design is to determine a suitable design point in the s-plane and use the point design technique. If the two feedback paths are combined, the block diagram shown in Fig. 14-31 is obtained.

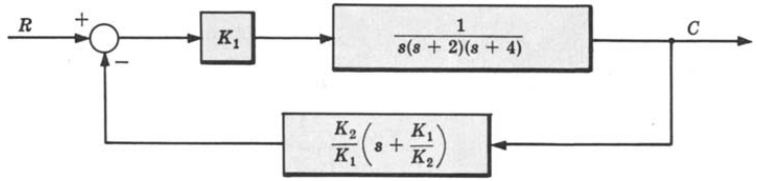


Fig. 14-31

For this configuration

$$GH = \frac{K_2(s + K_1/K_2)}{s(s+2)(s+4)}$$

The zero location at $s = -K_1/K_2$ appears in the feedback path and the gain factor is K_2. Thus for a fixed zero location (ratio of K_1/K_2), a root-locus for the system may be constructed as a function of K_2. The closed-loop transfer function will then contain three poles, but no zeros. Rough sketches of the root-locus (Fig. 14-32) reveal that if the ratio K_1/K_2 is set anywhere between 0 and 4, the closed-loop transfer function will probably contain two complex poles (if K_2 is large enough) and a real axis pole near the value of $-K_1/K_2$.

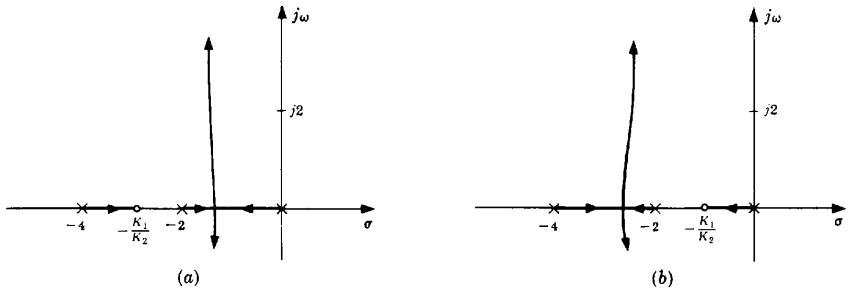


Fig. 14-32

A three-pole dominant configuration may then be appropriate for the design. A value of $\zeta = 0.5$ for the complex poles will satisfy the overshoot requirement. For $\zeta = 0.5$ and $p_r/\zeta\omega_n = 2$, Fig. 14-12 shows a normalized rise time $\omega_n T_r = 2.3$. Thus $T_r = 2.3/\omega_n < 1$ sec or $\omega_n > 2.3$ rad/sec. If $p_r/\zeta\omega_n$ turns out to be greater than 2, the rise time will be faster, and vice versa. In order to have a little margin in case $p_r/\zeta\omega_n$ is smaller than 2, let us choose $\omega_n = 2.6$. The design point in the s-plane is therefore $p_1 = -1.3 + j2.3$, corresponding to $\zeta = 0.5$ and $\omega_n = 2.6$.

From Fig. 14-33, the contribution of the poles at $s = 0$, -2, and -4 to $\arg GH(p_1)$ is $-233°$. The contribution of the zero must therefore be $-180° - (-233°) = 53°$ at p_1 to satisfy the angle criterion at p_1. The zero location is determined at $s = -3$ by drawing a line from p_1 to the real axis at 53°. With $K_1/K_2 = 3$, the gain factor at p_1 for GH is 7.5. Thus the design values are $K_2 = 7.5$ and $K_1 = 22.5$. The closed-loop real axis pole is to the left of, but near the zero located at $s = -3$. Therefore $p_r/\zeta\omega_n$ for this design is at least $3/1.3 = 2.3$.

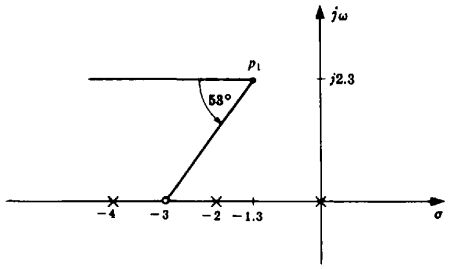


Fig. 14-33

14.17. For the discrete-time system with forward-loop transfer function

$$G_2 = \frac{K}{z(z-1)}$$

determine a feedback compensator that yields a closed-loop system with a deadbeat response.

For a deadbeat response (Section 10.8), the closed-loop transfer function must have all its poles at $z = 0$. Since poles cancelled by feedback zeros appear in the closed-loop transfer function, let H have a zero at $z = 0$. This eliminates the pole at $z = 0$ from the root-locus but it remains in the closed-loop transfer function.

For realizability, H must also have at least one pole. If we place the pole of H at $z = -1$, the resulting root-locus goes through $z = 0$, as shown in Fig. 14-34. Then, by setting $K = 1$, all the closed-loop poles are located at $z = 0$ and the system has a deadbeat response.

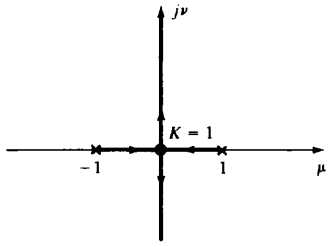


Fig. 14-34

Supplementary Problems

14.18. For the system with the open-loop transfer function $GH = K(s + a)/(s^2 - 1)(s + 5)$ determine K and a such that the closed-loop system has dominant poles with $\zeta = 0.5$ and $\omega_n = 2$. What is the percentage overshoot of the closed-loop system with these values of K and a?

14.19. Determine a suitable compensator for the system with the plant transfer function

$$G_2 = \frac{1}{s(s+1)(s+4)}$$

to satisfy the following specifications: (1) overshoot < 20%, (2) 10 to 90% rise time ≤ 1 sec, (3) gain margin ≥ 5.

14.20. Determine suitable compensation for the system with the plant transfer function $G_2 = 1/s(s + 4)^2$ to satisfy the following specifications: (1) overshoot < 20%, (2) velocity error constant $K_v \geq 10$.

14.21. For the system shown in the block diagram of Fig. 14-35, determine K_1 and K_2 such that the system has closed-loop poles at $s = -2 \pm j2$.

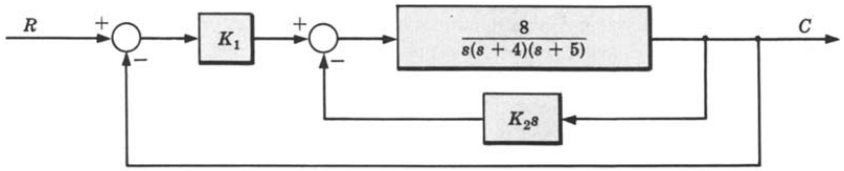


Fig. 14-35

14.22. Determine a value of K for the system with the open-loop transfer function $GH = K/s(s^2 + 6s + 25)$ such that the velocity error constant $K_v > 1$, the closed-loop step response has no overshoot, and the gain margin > 5.

14.23. Design a compensator for the system with the plant transfer function $G_2 = 63/s(s + 7)(s + 9)$ such that the velocity error constant $K_v > 30$, the overshoot is less than 20%, and the 10 to 90% rise time is less than 0.5 sec.

Answers to Supplementary Problems

14.18. $K = 11.25$, $a = 1.6$, overshoot = 38%; note that the system has a closed-loop zero at $s = -a = -1.6$.

14.19. $G_1 = 24(s + 1)/(s + 4)$

14.20. $G_1 = 24(s + 0.2)/(s + 0.03)$

14.21. $K_2 = 1$, $K_1 = 5$

14.22. $K = 28$

14.23. $G_1 = 3(s + 0.5)/(s + 0.05)$

Chapter 15

Bode Analysis

15.1 INTRODUCTION

The analysis of feedback control systems using the Bode method is equivalent to Nyquist analysis in that both techniques employ graphical representations of the open-loop frequency response function $GH(\omega)$, where $GH(\omega)$ refers to either a discrete-time or a continuous-time system. However, **Bode plots** consist of two graphs: the magnitude of $GH(\omega)$, and the phase angle of $GH(\omega)$, both plotted as a function of frequency ω. Logarithmic scales are usually used for the frequency axes and for $|GH(\omega)|$.

Bode plots clearly illustrate the relative stability of a system. In fact, gain and phase margins are often defined in terms of Bode plots (see Example 10.1). These measures of relative stability can be determined for a particular system with a minimum of computational effort using Bode plots, especially for those cases where experimental frequency response data are available.

15.2 LOGARITHMIC SCALES AND BODE PLOTS

Logarithmic scales are used for Bode plots because they considerably simplify their construction, manipulation, and interpretation.

A logarithmic scale is used for the ω-axes (abscissas) because the magnitude and phase angle may be graphed over a greater range of frequencies than with linear frequency axes, all frequencies being equally emphasized, and such graphs for continuous-time systems often result in straight lines (Section 15.4).

The magnitude $|P(\omega)|$ of any frequency response function $P(\omega)$ for any value of ω is plotted on a logarithmic scale in decibel (db) units, where

$$\text{db} = 20\log_{10}|P(\omega)| \tag{15.1}$$

[Also see Equation (*10.4*).]

EXAMPLE 15.1. If $|P(2)| \equiv |GH(2)| = 10$, the magnitude is $20\log_{10}10 = 20$ db.

Since the decibel is a logarithmic unit, the **db magnitude** of a frequency response function composed of a *product* of terms is equal to the *sum* of the db magnitudes of the individual terms. Thus, when the logarithmic scale is employed, the magnitude plot of a frequency response function expressible as a product of more than one term can be obtained by adding the individual db magnitude plots for each product term.

The *db magnitude versus log* ω plot is called the **Bode magnitude plot**, and the *phase angle versus log* ω plot is the **Bode phase angle plot**. The Bode magnitude plot is sometimes called the *log-modulus plot* in the literature.

EXAMPLE 15.2. The Bode magnitude plot for the continuous-time frequency response function

$$P(j\omega) = \frac{100[1 + j(\omega/10)]}{1 + j\omega}$$

may be obtained by adding the Bode magnitude plots for: 100, $1 + j(\omega/10)$, and $1/(1 + j\omega)$.

15.3 THE BODE FORM AND THE BODE GAIN FOR CONTINUOUS-TIME SYSTEMS

It is convenient to use the so-called *Bode form* of a continuous-time frequency response function when using Bode plots for analysis and design because of the asymptotic approximations in Section 15.4.

The **Bode form** for the function

$$\frac{K(j\omega + z_1)(j\omega + z_2)\cdots(j\omega + z_m)}{(j\omega)^l(j\omega + p_1)(j\omega + p_2)\cdots(j\omega + p_n)}$$

where l is a nonnegative integer, is obtained by factoring out all z_i and p_i and rearranging it in the form

$$\frac{\left[K\prod_{i=1}^{m} z_i \Big/ \prod_{i=1}^{n} p_i\right](1 + j\omega/z_1)(1 + j\omega/z_2)\cdots(1 + j\omega/z_m)}{(j\omega)^l(1 + j\omega/p_1)(1 + j\omega/p_2)\cdots(1 + j\omega/p_n)} \tag{15.2}$$

The **Bode gain** K_B is defined as the coefficient of the numerator in Equation (*15.2*):

$$K_B \equiv \frac{K\prod_{i=1}^{m} z_i}{\prod_{i=1}^{n} p_i} \tag{15.3}$$

15.4 BODE PLOTS OF SIMPLE CONTINUOUS-TIME FREQUENCY RESPONSE FUNCTIONS AND THEIR ASYMPTOTIC APPROXIMATIONS

The constant K_B has a magnitude $|K_B|$, a phase angle of 0° if K_B is positive, and $-180°$ if K_B is negative. Therefore the Bode plots for K_B are simply horizontal straight lines as shown in Figs. 15-1 and 15-2.

Fig. 15-1

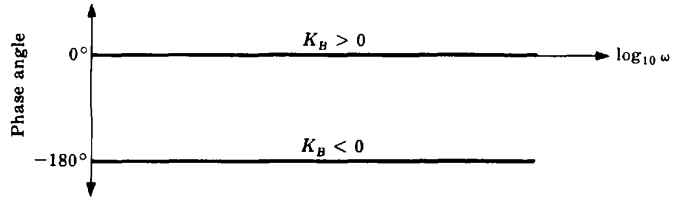


Fig. 15-2

The frequency response function (or sinusoidal transfer function) for a *pole of order l at the origin* is

$$\frac{1}{(j\omega)^l} \tag{15.4}$$

The bode plots for this function are straight lines, as shown in Figs. 15-3 and 15-4.

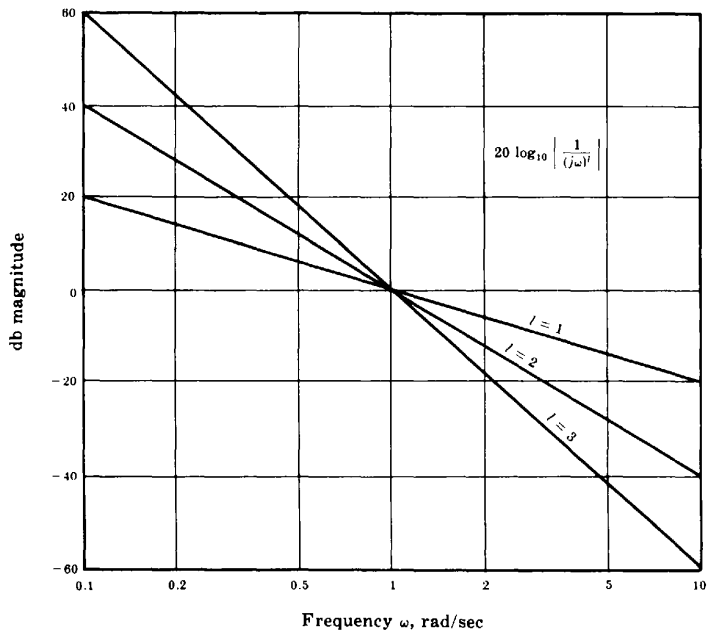


Fig. 15-3

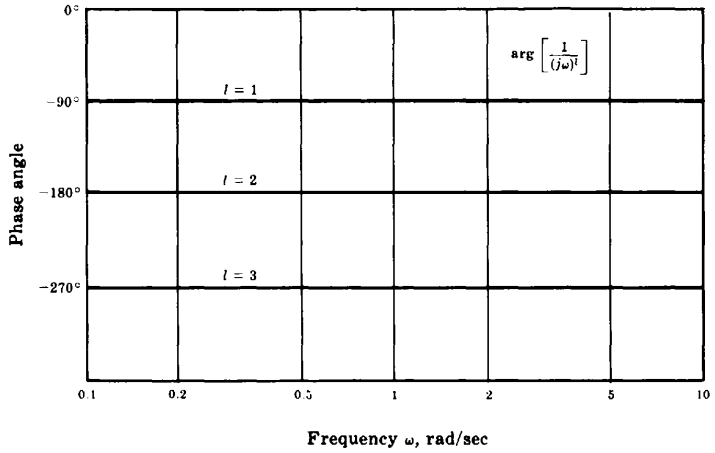


Fig. 15-4

For a *zero of order l at the origin*,

$$(j\omega)^l. \tag{15.5}$$

the Bode plots are the reflections about the 0-db and 0° lines of Figs. 15-3 and 15-4, as shown in Figs. 15-5 and 15-6.

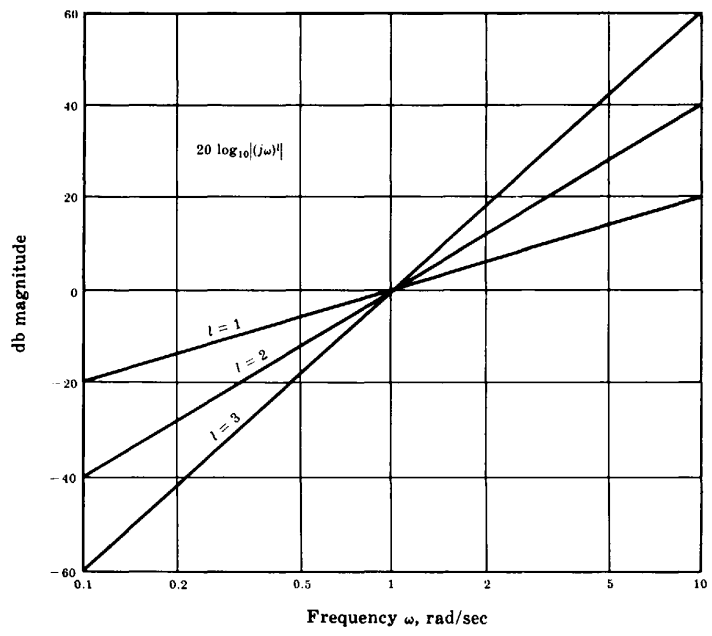


Fig. 15-5

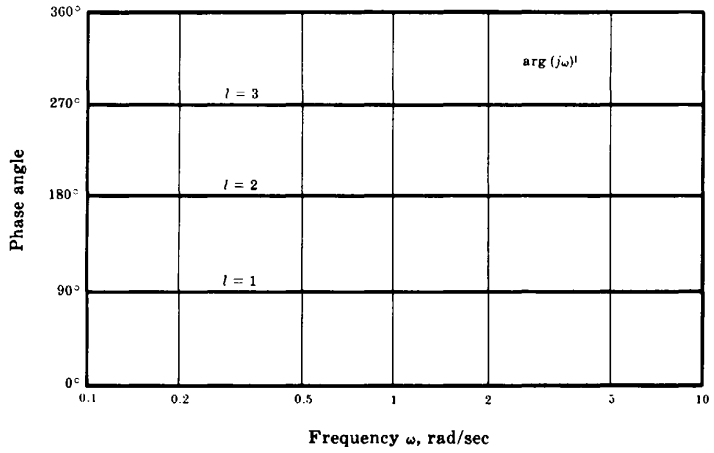


Fig. 15-6

Consider the *single-pole* transfer function $p/(s+p)$, $p > 0$. The Bode plots for its frequency response function

$$\frac{1}{1 + j\omega/p} \tag{15.6}$$

are given in Figs. 15-7 and 15-8. Note that the logarithmic frequency scale is normalized in terms of p.

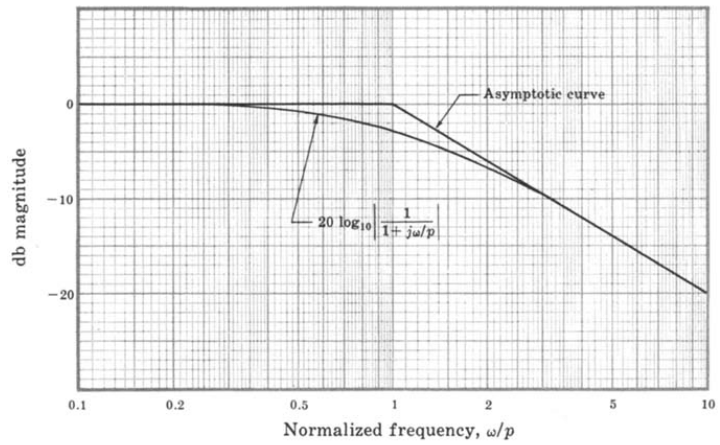


Fig. 15-7

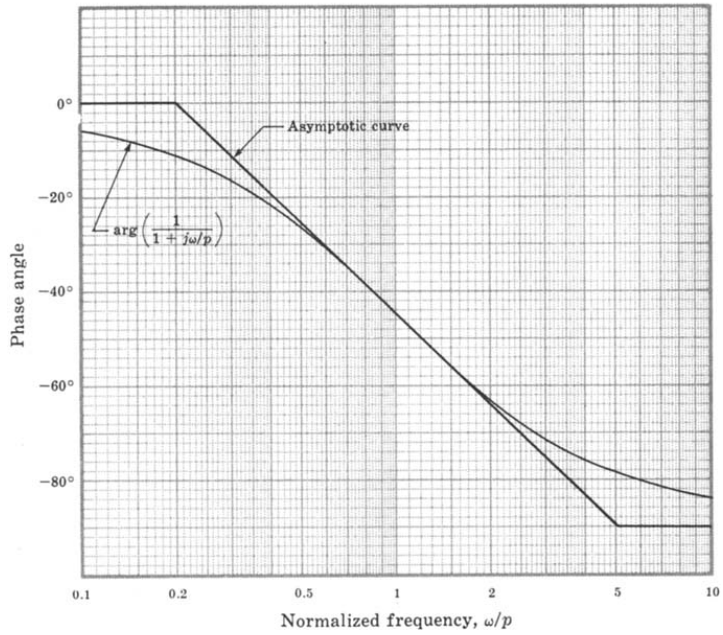


Fig. 15-8

To determine the *asymptotic approximations* for these Bode plots, we see that for $\omega/p \ll 1$, or $\omega \ll p$,

$$20 \log_{10}\left|\frac{1}{1+j\omega/p}\right| \cong 20 \log_{10} 1 = 0 \text{ db}$$

and for $\omega/p \gg 1$, or $\omega \gg p$,

$$20 \log_{10}\left|\frac{1}{1+j\omega/p}\right| \cong 20 \log_{10}\left|\frac{1}{j\omega/p}\right| = -20 \log_{10}\left(\frac{\omega}{p}\right)$$

Therefore the Bode magnitude plot asymptotically approaches a horizontal straight line at 0 db as ω/p approaches zero and $-20 \log_{10}(\omega/p)$ as ω/p approaches infinity (Fig. 15-7). Note that this high-frequency asymptote is a straight line with a slope of -20 db/decade, or -6 db/octave when plotted on a logarithmic frequency scale as shown in Fig. 15-7. The two asymptotes intersect at the **corner frequency** $\omega = p$ rad/sec. To determine the phase angle asymptote, we see that for $\omega/p \ll 1$, or $\omega \ll p$,

$$\arg\left(\frac{1}{1+j\omega/p}\right) = -\tan^{-1}\left(\frac{\omega}{p}\right)\Big|_{\omega \ll p} \cong 0°$$

and for $\omega/p \gg 1$, or $\omega \gg p$,

$$\arg\left(\frac{1}{1+j\omega/p}\right) = -\tan^{-1}\left(\frac{\omega}{p}\right)\Big|_{\omega \gg p} \cong -90°$$

Thus the Bode phase angle plot asymptotically approaches 0° as ω/p approaches zero, and $-90°$ as ω/p approaches infinity, as shown in Fig. 15-8. A negative-slope straight-line asymptote can be used to join the 0° asymptote and the $-90°$ asymptote by drawing a line from the 0° asymptote at $\omega = p/5$ to the $-90°$ asymptote at $\omega = 5p$. Note that it is tangent to the exact curves at $\omega = p$.

The *errors* introduced by these asymptotic approximations are shown in Table 15-1 for the single-pole transfer function at various frequencies.

Table 15-1. Asymptotic Errors for $\frac{1}{1+j\omega/p}$

ω	$p/5$	$p/2$	p	$2p$	$5p$
Magnitude error (db)	−0.17	−0.96	−3	−0.96	−0.17
Phase angle error	−11.3°	−0.8°	0°	+0.8°	+11.3°

The Bode plots and their asymptotic approximations for the *single-zero* frequency response function

$$1 + \frac{j\omega}{z_1} \tag{15.7}$$

are shown in Figs. 15-9 and 15-10.

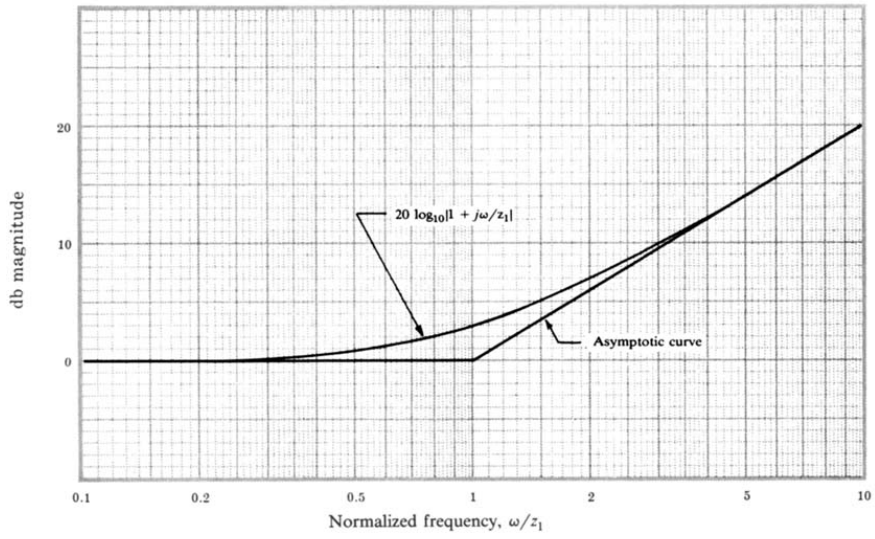


Fig. 15-9

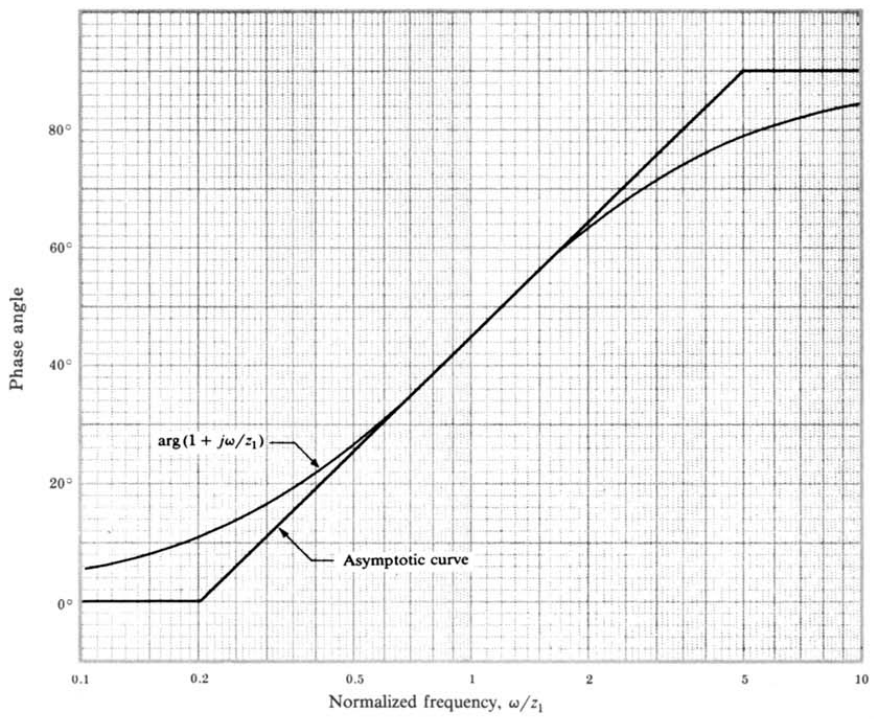


Fig. 15-10

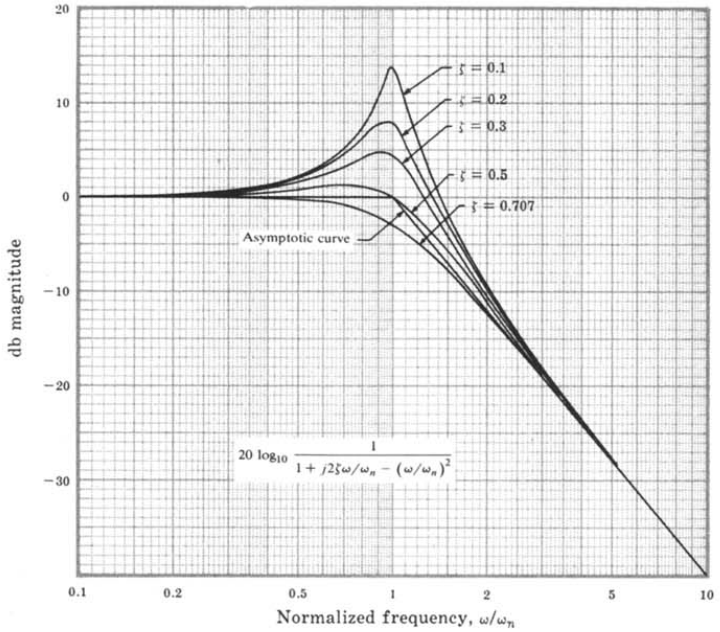


Fig. 15-11

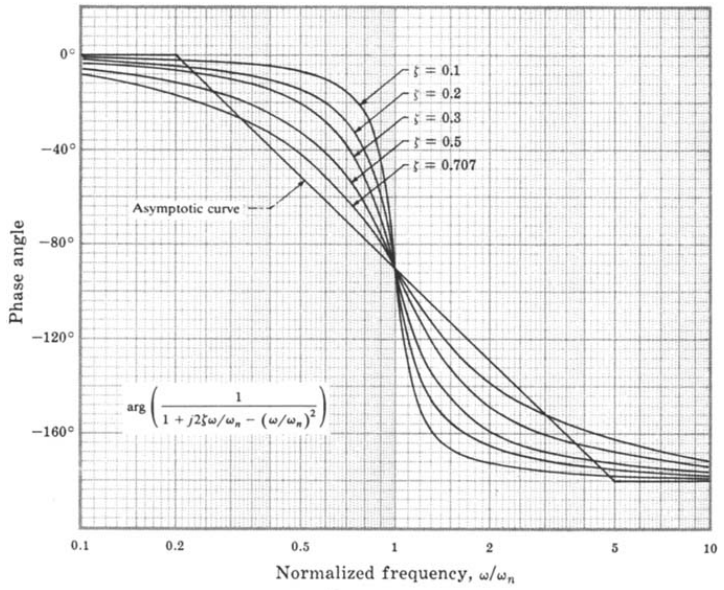


Fig. 15-12

The Bode plots and their asymptotic approximations for the second-order frequency response function with *complex poles*,

$$\frac{1}{1+j2\zeta\omega/\omega_n-(\omega/\omega_n)^2} \qquad 0\le\zeta\le 1 \tag{15.8}$$

are shown in Figs. 15-11 and 15-12. Note that the damping ratio ζ is a parameter on these graphs.

The magnitude asymptote shown in Fig. 15-11 has a corner frequency at $\omega=\omega_n$ and a high-frequency slope twice that of the asymptote for the single-pole case of Fig. 15-7. The phase angle asymptote is similar to that of Fig. 15-8 except that the high-frequency portion is at $-180°$ instead of $-90°$ and the point of tangency, or inflection, is at $-90°$.

The Bode plots for a pair of *complex zeros* are the reflections about the 0 db and 0° lines of those for the complex poles.

15.5 CONSTRUCTION OF BODE PLOTS FOR CONTINUOUS-TIME SYSTEMS

Bode plots of continuous-time frequency response functions can be constructed by summing the magnitude and phase angle contributions of each pole and zero (or pairs of complex poles and zeros). The asymptotic approximations of these plots are often sufficient. If more accurate plots are desired, many software packages are available for rapidly accomplishing this task.

For the general open-loop frequency response function

$$GH(j\omega)=\frac{K_B(1+j\omega/z_1)(1+j\omega/z_2)\cdots(1+j\omega/z_m)}{(j\omega)^l(1+j\omega/p_1)(1+j\omega/p_2)\cdots(1+j\omega/p_n)} \tag{15.9}$$

where l is a positive integer or zero, the magnitude and phase angle are given by

$$20\log_{10}|GH(j\omega)|=20\log_{10}|K_B|+20\log_{10}\left|1+\frac{j\omega}{z_1}\right|+\cdots+20\log_{10}\left|1+\frac{j\omega}{z_m}\right|$$
$$+20\log_{10}\frac{1}{|(j\omega)^l|}+20\log_{10}\frac{1}{|1+j\omega/p_1|}+\cdots+20\log_{10}\frac{1}{|1+j\omega/p_n|} \tag{15.10}$$

and

$$\arg GH(j\omega) = \arg K_B + \arg\left(1 + \frac{j\omega}{z_1}\right) + \cdots + \arg\left(1 + \frac{j\omega}{z_m}\right) + \arg\left(\frac{1}{(j\omega)^l}\right) + \arg\left(\frac{1}{1 + j\omega/p_1}\right) + \cdots + \arg\left(\frac{1}{1 + j\omega/p_n}\right) \qquad (15.11)$$

The Bode plots for each of the terms in Equations (*15.10*) and (*15.11*) were given in Figs. 15-1 to 15-12. If $GH(j\omega)$ has complex poles or zeros, terms having a form similar to Equation (*15.8*) are simply added to Equations (*15.10*) and (*15.11*). The construction procedure is best illustrated by an example.

EXAMPLE 15.3. The asymptotic Bode plots for the frequency response function

$$GH(j\omega) = \frac{10(1 + j\omega)}{(j\omega)^2[1 + j\omega/4 - (\omega/4)^2]}$$

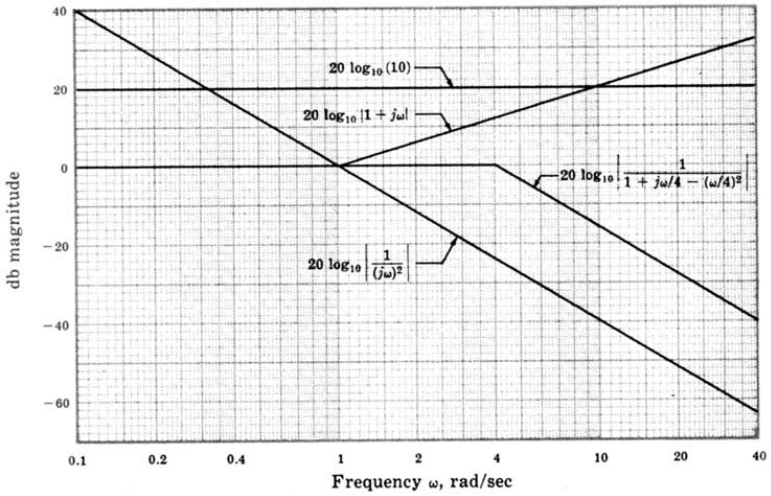


Fig. 15-13

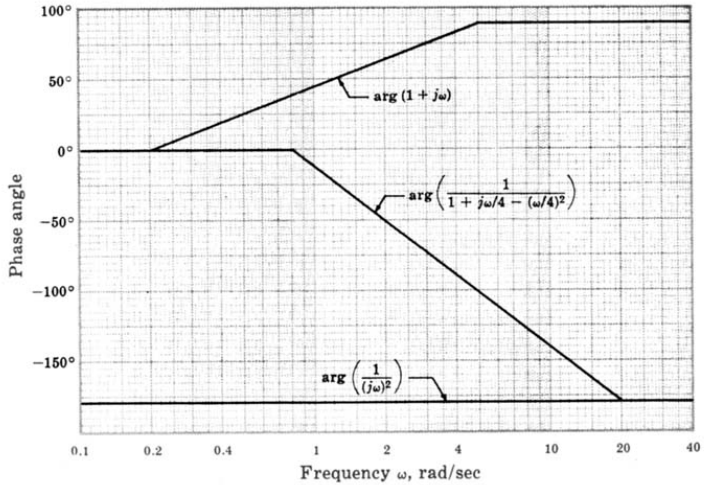


Fig. 15-14

are constructed using Equations (*15.10*) and (*15.11*):

$$20\log_{10}|GH(j\omega)| = 20\log_{10}10 + 20\log_{10}|1 + j\omega| + 20\log_{10}\left|\frac{1}{(j\omega)^2}\right| + 20\log_{10}\left|\frac{1}{1 + j\omega/4 - (\omega/4)^2}\right|$$

$$\arg GH(j\omega) = \arg(1 + j\omega) + \arg\left(1/(j\omega)^2\right) + \arg\left(\frac{1}{1 + j\omega/4 - (\omega/4)^2}\right)$$

The graphs for each of the terms in these equations are obtained from Figs. 15-1 to 15-12 and are shown in Figs. 15-13 and 15-14. The asymptotic Bode plots for $GH(j\omega)$ are obtained by adding these curves, as shown in Figs. 15-15 and 15-16, where computer-generated Bode plots for the frequency response function are also given for comparison with the asymptotic approximations.

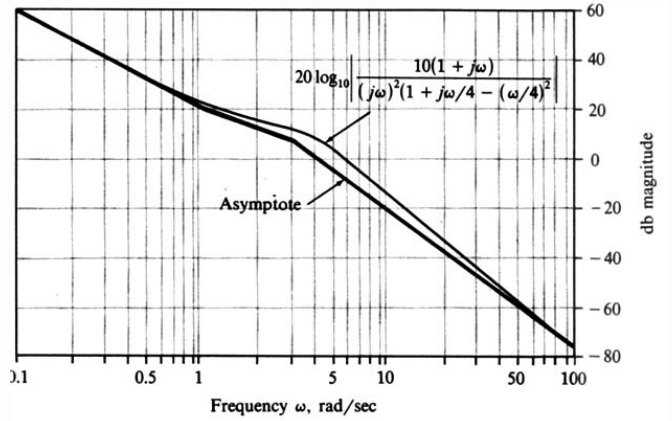


Fig. 15-15

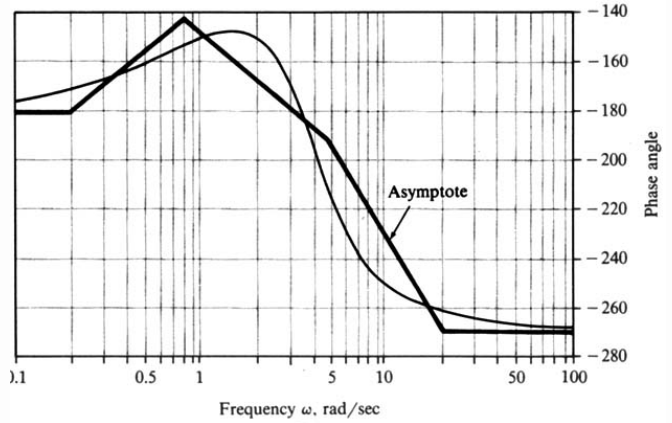


Fig. 15-16

15.6 BODE PLOTS OF DISCRETE-TIME FREQUENCY RESPONSE FUNCTIONS

The factored form for the general open-loop discrete-time frequency response function is

$$GH(e^{j\omega T}) = \frac{K(e^{j\omega T} + z_1)(e^{j\omega T} + z_2)\cdots(e^{j\omega T} + z_m)}{(e^{j\omega T} + p_1)(e^{j\omega T} + p_2)\cdots(e^{j\omega T} + p_n)} \qquad (15.12)$$

Simple asymptotic approximations, similar to those in Section 15.4, do not exist for the individual terms in Equation (*15.12*). Thus there is no particular advantage to a *Bode form* of the type in Equation (*15.2*) for discrete-time systems. In general, computers provide the most convenient way to generate Bode plots for discrete-time systems and several software packages exist to accomplish this task.

For the general open-loop frequency response function Equation (*15.12*), the magnitude and phase angle are given by

$$\begin{aligned} 20\log_{10}|GH(e^{j\omega T})| &= 20\log_{10}|K| \\ &+ 20\log_{10}|e^{j\omega T}+z_1| + \cdots + 20\log_{10}|e^{j\omega T}+z_m| \\ &+ 20\log_{10}\frac{1}{|e^{j\omega T}+p_1|} + \cdots + 20\log_{10}\frac{1}{|e^{j\omega T}+p_n|} \end{aligned} \qquad (15.13)$$

and

$$\begin{aligned} \arg GH(e^{j\omega T}) &= \arg K + \arg(e^{j\omega T}+z_1) + \cdots \\ &+ \arg(e^{j\omega T}+z_m) + \arg\frac{1}{(e^{j\omega T}+p_1)} + \cdots + \arg\left(\frac{1}{e^{j\omega T}+p_n}\right) \end{aligned} \qquad (15.14)$$

It is important to note that both the magnitude and phase angle of discrete-time frequency response functions are periodic in the real angular frequency variable ω. This is true since

$$e^{j\omega T} = e^{j(\omega+2k\pi/T)T} = e^{j\omega T}e^{j2k\pi}$$

thus $e^{j\omega T}$ is periodic in the frequency domain with period $2\pi/T$. Every term in both the magnitude and phase angle is thus periodic. It is therefore only necessary to generate Bode plots over the angular range $-\pi \le \omega T \le \pi$ radians; and the magnitude and phase angle are typically plotted as a function of the angle ωT rather than angular frequency ω.

Another useful property of a discrete-time frequency response function is that the magnitude is an even function of the frequency ω (and ωT) and the phase angle is an odd function of ω (and ωT).

EXAMPLE 15.4. The Bode plots for the discrete-time frequency response function

$$GH(e^{j\omega T}) = \frac{\frac{1}{100}(e^{j\omega T}+1)^2}{(e^{j\omega T}-1)(e^{j\omega T}+\frac{1}{3})(e^{j\omega T}+\frac{1}{2})}$$

are shown in Figs. 15-17 and 15-18.

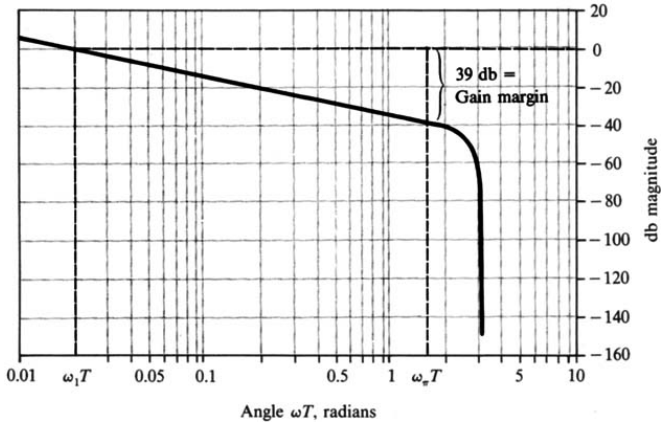


Fig. 15-17

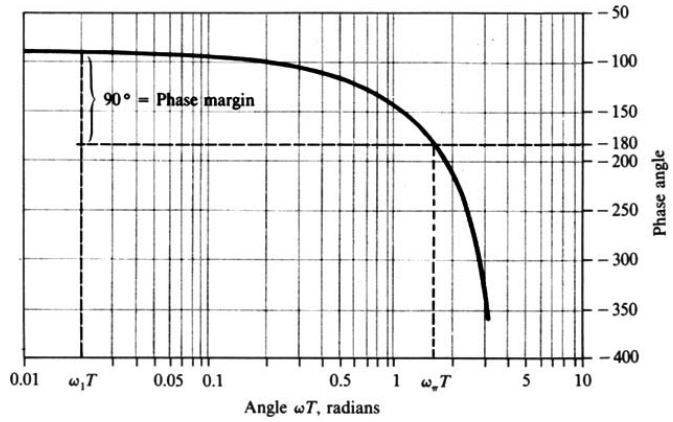


Fig. 15-18

15.7 RELATIVE STABILITY

The relative stability indicators "gain margin" and "phase margin" for both discrete-time and continuous-time systems are defined in terms of the system open-loop frequency response in Section 10.4. Consequently these parameters are easily determined from the Bode plots of $GH(\omega)$ as illustrated in Example 10.1, and in Example 15.4 above. Since 0 db corresponds to a magnitude of 1, the **gain margin** is the number of decibels that $|GH(\omega)|$ is *below* 0 db at the phase crossover frequency ω_π ($\arg GH(\omega_\pi) = -180°$). The **phase margin** is the number of degrees $\arg GH(\omega)$ is above $-180°$ at the gain crossover frequency ω_1 ($|GH(\omega_1)| = 1$). Computer-generated Bode plots should be used to accurately determine ω_π, ω_1 and the gain and phase margins.

In most cases positive gain and phase margins, as defined above, will ensure stability of the closed-loop system. However, a Nyquist Stability Plot (Chapter 11) may be sketched, or one of the methods of Chapter 5 can be used to verify the absolute stability of the system.

EXAMPLE 15.5. The continuous-time system whose Bode plots are shown in Fig. 15-19 has a gain margin of 8 db and a phase margin of 40°.

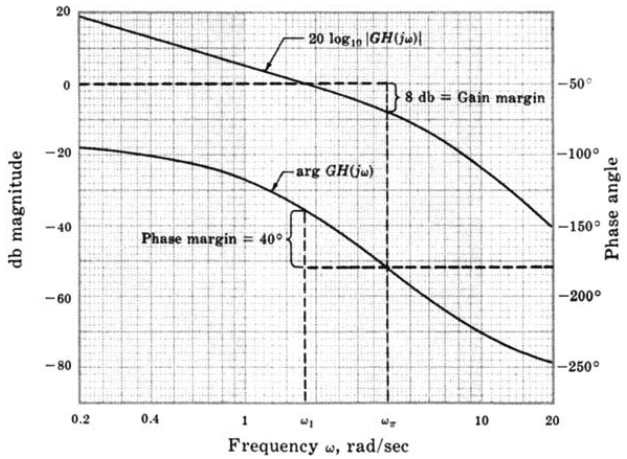


Fig. 15-19

EXAMPLE 15.6. For the system in Example 15.4, the gain margin is 39 db, the angle at the phase crossover frequency ω_π is $\omega_\pi T = 1.57$ rad, the phase margin is 90°, and the angle at the gain crossover frequency ω_1 is $\omega_1 T = 0.02$ rad, all as illustrated in Figures 15-17 and 15-18.

15.8 CLOSED-LOOP FREQUENCY RESPONSE

Although there is no straightforward method for plotting the closed-loop frequency response $(C/R)(\omega)$ from Bode plots of $GH(\omega)$, it may be approximated in the following manner, for both continuous and discrete-time control systems. The closed-loop frequency response is given by

$$\frac{C}{R}(\omega) = \frac{G(\omega)}{1 + GH(\omega)}$$

If $|GH(\omega)| \gg 1$,

$$\left.\frac{C}{R}(\omega)\right|_{|GH(\omega)|\gg 1} \cong \frac{G(\omega)}{GH(\omega)} = \frac{1}{H(\omega)}$$

If $|GH(\omega)| \ll 1$,

$$\left.\frac{C}{R}(\omega)\right|_{|GH(\omega)|\ll 1} \cong G(\omega)$$

The open-loop frequency response of most systems is characterized by high gain for low frequencies and decreasing gain for higher frequencies, due to the usual excess of poles over zeros. Thus the closed-loop frequency response *for a unity feedback system* ($H = 1$) is approximated by a magnitude of 1 (0 db) and phase angle of 0° for frequencies below the gain crossover frequency ω_1. For frequencies above ω_1, the closed-loop frequency response may be approximated by the magnitude and phase angle of $G(\omega)$. *An approximate closed-loop bandwidth for many systems is the gain crossover frequency* ω_1 (See Example 12.7.)

EXAMPLE 15.7. The open-loop Bode magnitude plot and approximate closed-loop Bode magnitude plot for the continuous-time unity feedback system represented by $G(j\omega) = 10/j\omega(1 + j\omega)$ are shown in Fig. 15-20.

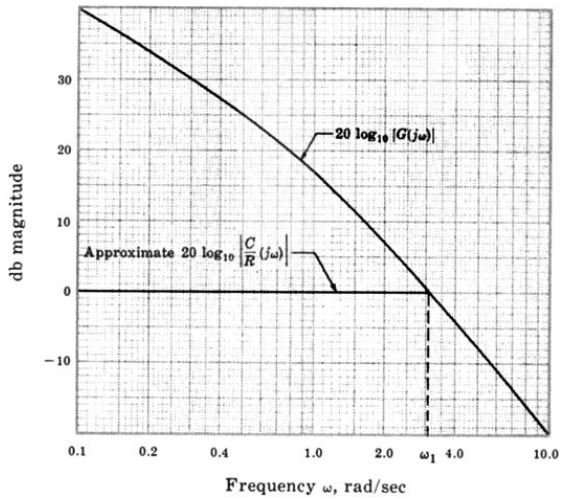


Fig. 15-20

15.9 BODE ANALYSIS OF DISCRETE-TIME SYSTEMS USING THE *w*-TRANSFORM

The *w*-transform discussed in Section 10.7 can be used in the Bode analysis of discrete-time systems. The algorithm for Bode analysis using the *w*-transform is:

1. Substitute $(1 + w)/(1 - w)$ for z in the open-loop transfer function $GH(z)$:
$$GH(z)\big|_{z=\frac{1+w}{1-w}} \equiv GH'(w)$$
2. Let $w = j\omega_w$ and generate Bode plots for $GH'(j\omega_w)$, using the methods of Sections 15.3 through 15.5.
3. Analyze the relative stability of the system in the *w*-plane by determining the gain and phase margins, the gain and phase crossover frequencies, the closed-loop frequency response, the bandwidth, and/or any other frequency-related characteristics of interest.
4. Transform the critical frequencies determined in step 3 to the frequency domain of the *z*-plane using the transformation $\omega T = 2\tan^{-1}\omega_w$.

EXAMPLE 15.8. The open-loop transfer function

$$GH(z) = \frac{\frac{1}{100}(z+1)^2}{(z-1)(z+\frac{1}{3})(z+\frac{1}{2})}$$

is transformed into the *w*-domain by letting

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

which yields

$$GH'(w) = \frac{-\frac{6}{100}(w-1)}{w(w+2)(w+3)}$$

Note, in particular, that the minus sign contributes −180° of phase angle, and the zero at +1 contributes +90° at $\omega_w = 0$°. The Bode plots of $GH'(j\omega_w)$ are shown in Figs. 15-21 and 15-22.

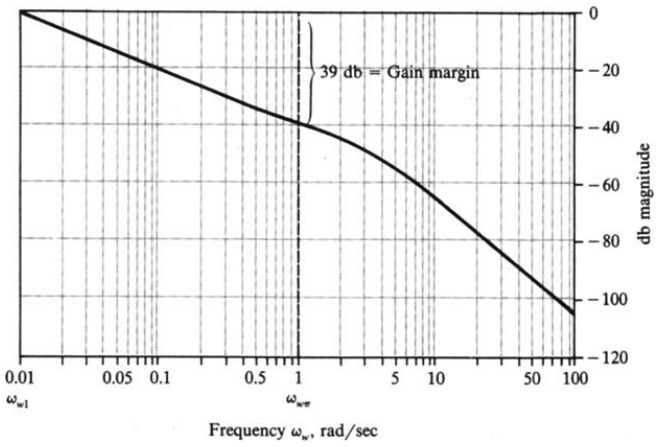


Fig. 15-21

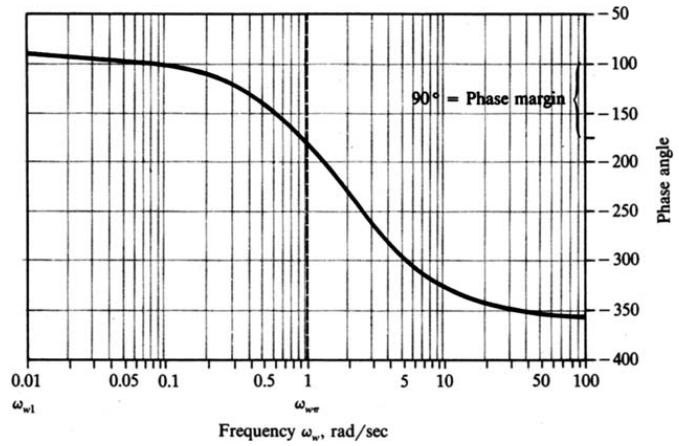


Fig. 15-22

EXAMPLE 15.9. From the Bode plots of Example 15.8, the gain margin in the w-domain is 39 db and the phase crossover frequency is $\omega_{w\pi} = 1$ rad/sec. Transforming back to the z-domain, the phase crossover frequency ω_π is obtained from

$$\omega_\pi T = 2\tan^{-1}\omega_{w\pi} = 1.57 \text{ rad}$$

Compare these results with those of Example 15.6, which are the same.

EXAMPLE 15.10. From the Bode plots of Example 15.8, the phase margin is 90° and the gain crossover frequency is $\omega_{w1} = 0.01$ rad/sec. Transforming to the z-domain, the gain crossover frequency ω_1 is obtained from

$$\omega_1 T = 2\tan^{-1}\omega_{w1} = 0.02 \text{ rad}$$

Compare these results with those of Example 15.6, which are the same.

With the wide availability of software for control systems analysis, use of the w-transform for Bode *analysis* of discrete-time systems is usually unnecessary. However, for *design* by analysis, as discussed in Chapter 16 where insight gained from continuous-time system design techniques is transferred to discrete-time system design, the w-transform can be a very useful tool.

Solved Problems

LOGARITHMIC SCALES

15.1. Express the following quantities in decibel (db) units: (*a*) 2, (*b*) 4, (*c*) 8, (*d*) 20, (*e*) 25, (*f*) 140.

From Equation (*15.1*),

$$\text{db}_a = 20\log_{10}2 = 20(0.301) = 6.02 \qquad \text{db}_d = 20\log_{10}20 = 20(1.301) = 26.02$$
$$\text{db}_b = 20\log_{10}4 = 20(0.602) = 12.04 \qquad \text{db}_e = 20\log_{10}25 = 20(1.398) = 27.96$$
$$\text{db}_c = 20\log_{10}8 = 20(0.903) = 18.06 \qquad \text{db}_f = 20\log_{10}140 = 20(2.146) = 42.92$$

Note that since $4 = 2 \times 2$, then for part (*b*) we have

$$20\log_{10}4 = 20\log_{10}2 + 20\log_{10}2 = 12.04$$

and since $8 = 2 \times 4$, then for part (*c*) we have

$$20\log_{10}8 = 20\log_{10}2 + 20\log_{10}4 = 6.02 + 12.04 = 18.06$$

THE BODE FORM AND THE BODE GAIN FOR CONTINUOUS-TIME SYSTEMS

15.2. Determine the Bode form and the Bode gain for the transfer function

$$GH = \frac{K(s+2)}{s^2(s+4)(s+6)}$$

Factoring 2 from the numerator, 4 and 6 from the denominator and putting $s = j\omega$ results in the Bode form

$$GH(j\omega) = \frac{(K/12)(1+j\omega/2)}{(j\omega)^2(1+j\omega/4)(1+j\omega/6)}$$

The Bode gain is $K_B = K/12$.

15.3. When is the Bode gain equal to the d.c. gain (zero frequency magnitude) of a transfer function?

The Bode gain is equal to the d.c. gain of any transfer function with no poles or zeros at the origin [$l = 0$ in Equation (*15.2*)].

BODE PLOTS OF SIMPLE FREQUENCY RESPONSE FUNCTIONS

15.4. Prove that the Bode Magnitude plot for $(j\omega)^l$ is a straight line.

The Bode magnitude plot for $(j\omega)^l$ is a plot of $20\log_{10}\omega^l$ versus $\log_{10}\omega$. Thus

$$\text{slope} = \frac{d(20\log_{10}\omega^l)}{d(\log_{10}\omega)} = \frac{20l\,d(\log_{10}\omega)}{d(\log_{10}\omega)} = 20l$$

Since the slope is constant for any l, the Bode magnitude plot is a straight line.

15.5. Determine: (1) the conditions under which the Bode magnitude plot for a pair of complex poles has a peak at a nonzero, finite value of ω; and (2) the frequency at which the peak occurs.

The Bode magnitude is given by

$$20\log_{10}\left|\frac{1}{1+j2\zeta\omega/\omega_n-(\omega/\omega_n)^2}\right|$$

Since the logarithm is a monotonically increasing function, the magnitude in decibels has a peak (maximum) if and only if the magnitude itself is maximum. The magnitude squared, which is maximum when the magnitude is maximum, is

$$\frac{1}{\left[1-(\omega/\omega_n)^2\right]^2+4(\zeta\omega/\omega_n)^2}$$

Taking the derivative of this function and setting it equal to zero yields

$$\frac{(4\omega/\omega_n^2)\left[1-(\omega/\omega_n)^2\right]-8\zeta^2\omega/\omega_n^2}{\left\{\left[1-(\omega/\omega_n)^2\right]^2+4(\zeta\omega/\omega_n)^2\right\}^2}=0$$

or

$$1-\left(\frac{\omega}{\omega_n}\right)^2-2\zeta^2=0$$

and the frequency at the peak is $\omega=\omega_n\sqrt{1-2\zeta^2}$. Since ω must be real, by definition, the magnitude has a peak at a nonzero value ω only if $1-2\zeta^2>0$ or $\zeta<1/\sqrt{2}=0.707$. For $\zeta\geq 0.707$, the Bode magnitude is monotonically decreasing.

CONSTRUCTION OF BODE PLOTS FOR CONTINUOUS-TIME SYSTEMS

15.6. Construct the asymptotic Bode plots for the frequency response function

$$GH(j\omega)=\frac{1+j\omega/2-(\omega/2)^2}{j\omega(1+j\omega/0.5)(1+j\omega/4)}$$

The asymptotic Bode plots are determined by summing the graphs of the asymptotic representations for each of the terms of $GH(j\omega)$, as in equations (*15.10*) and (*15.11*). The asymptotes for each of these terms are shown in Figs. 15-23 and 15-24 and the asymptotic Bode plots for $GH(j\omega)$ in Figs. 15-25 and 15-26. The exact Bode plots generated by computer are shown for comparison.

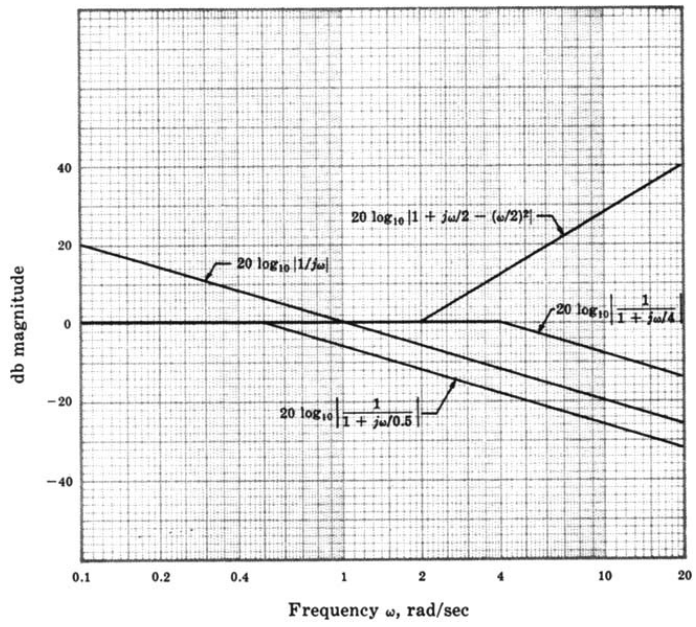


Fig. 15-23

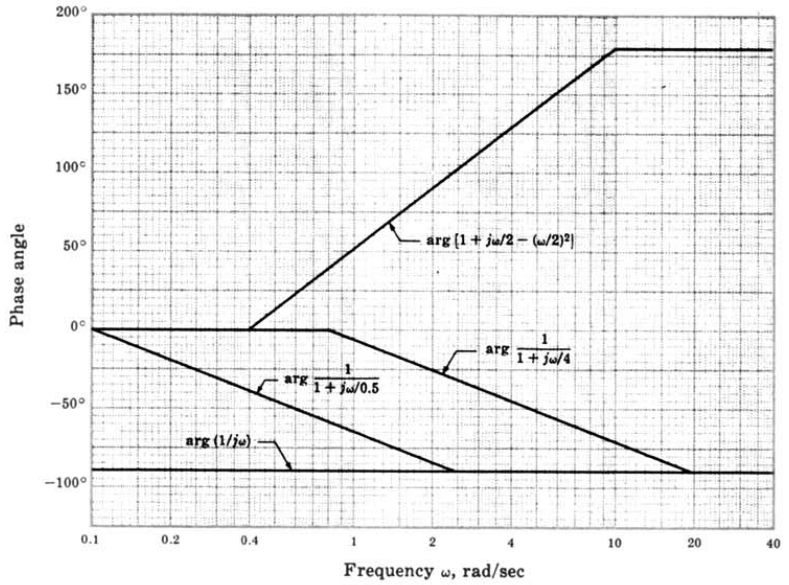


Fig. 15-24

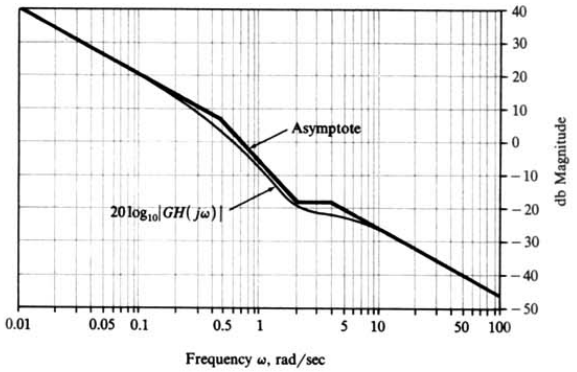


Fig. 15-25

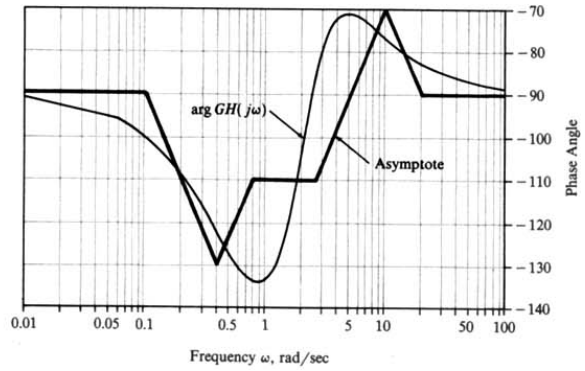


Fig. 15-26

15.7. Construct Bode plots for the frequency response function

$$GH(j\omega) = \frac{2}{j\omega(1 + j\omega/2)(1 + j\omega/5)}$$

The asymptotic Bode plots are constructed by summing the asymptotic plots for each term of $GH(j\omega)$, as in Equation (*15.10*) and (*15.11*), and are shown in Figs. 15-27 and 15-28. More accurate curves determined numerically by computer are also plotted for comparison.

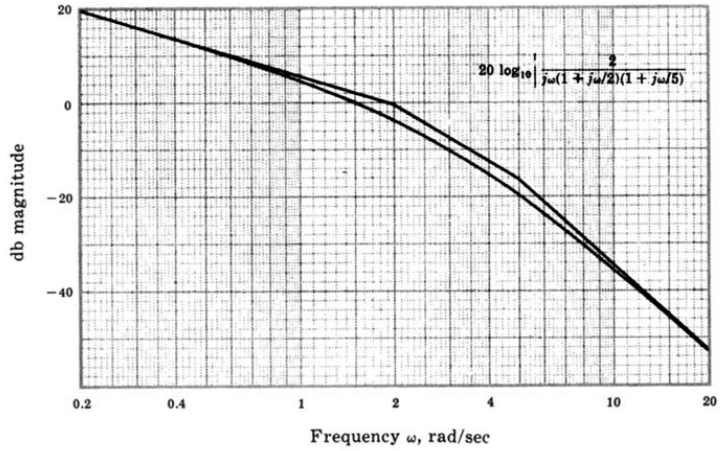


Fig. 15-27

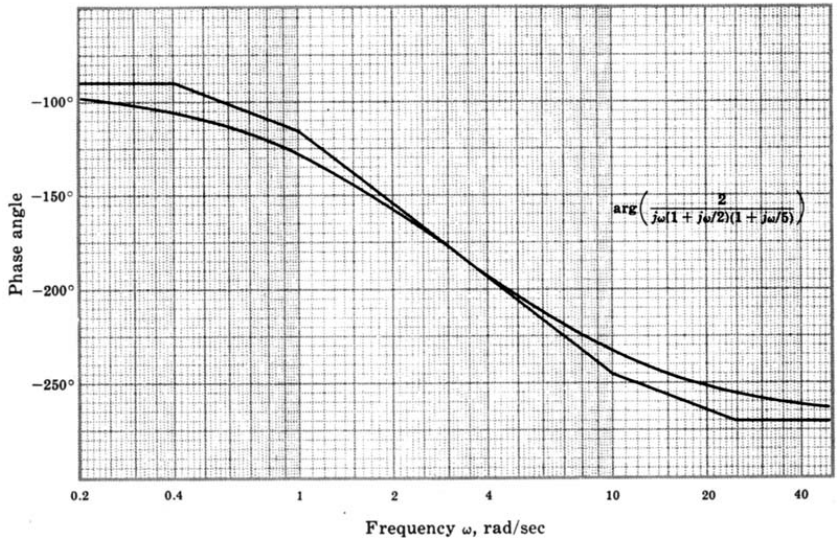


Fig. 15-28

15.8. Construct the Bode plots for the open-loop transfer function $GH = 2(s + 2)/(s^2 - 1)$.

With $s = j\omega$, the Bode form for this transfer function is

$$GH(j\omega) = \frac{-4(1 + j\omega/2)}{(1 + j\omega)(1 - j\omega)}$$

This function has a right-half plane pole [due to the term $1/(1 - j\omega)$] which is not one of the standard functions introduced in Section 15.4. However, this function has the same magnitude as $1/(1 + j\omega)$ and the same phase angle as $1 + j\omega$. Thus for a function of the form $1/(1 - j\omega/p)$, the magnitude can be determined from Fig. 15-7 and the phase angle from Fig. 15-10. For this problem the phase angle contributions from the terms $1/(1 + j\omega)$ and $1/(1 - j\omega)$ cancel each other. The asymptotes for the Bode magnitude plot are shown in Fig. 15-29 along with a more accurate Bode magnitude plot. The Bode phase angle is determined solely from $\arg K_B = \arg(-4) - 180°$ and the zero at $\omega = 2$, as shown in Fig. 15-30.

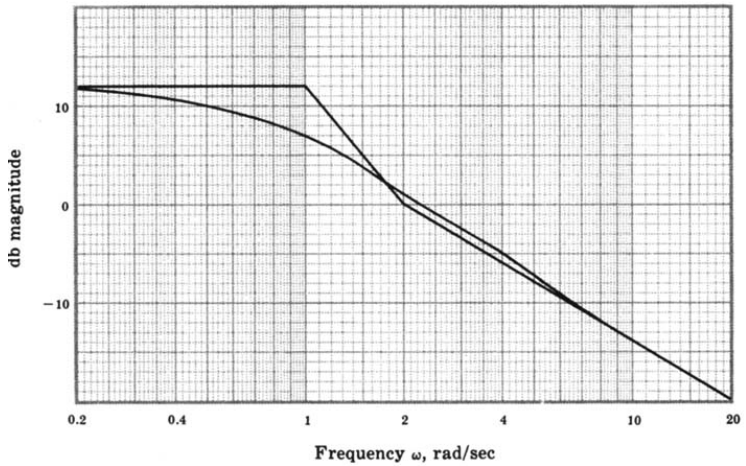


Fig. 15-29

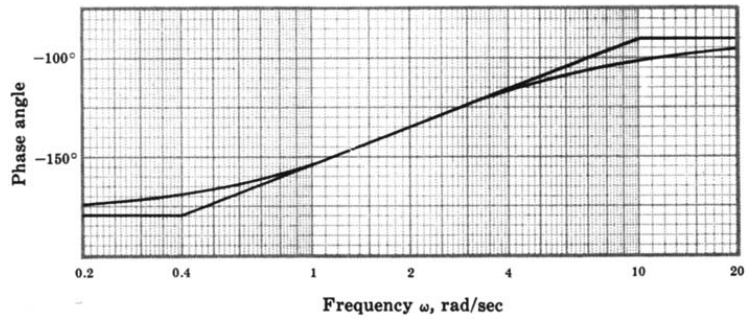


Fig. 15-30

RELATIVE STABILITY

15.9. For the system with the open-loop transfer function of Problem 15.6, find ω_1, ω_π, the gain margin, and the phase margin.

Using the exact magnitude curve shown in Fig. 15-25, the gain crossover frequency is $\omega_1 = 0.62$. The phase crossover frequency ω_π is indeterminate because $\arg GH(j\omega)$ never crosses $-180°$. (See Fig. 15-26.) $\text{Arg}\, GH(j\omega_1) = \arg GH(j0.62)$ is $-129°$. Hence the phase margin is $-129° + 180° = 51°$. Since ω_π is indeterminate, the gain margin is also indeterminate.

15.10. Determine the gain and phase margins for the systems with the open-loop frequency response function of Problem 15.7.

From Fig. 15-27, $\omega_1 = 1.5$; and from Fig. 15-28, $\arg GH(j\omega_1) = -144°$. Therefore the phase margin is $180° - 144° = 36°$. From Fig. 15-28, $\omega_\pi = 3.2$; and the gain margin is read from Fig. 15-27 as $-20\log_{10}|GH(j\omega_\pi)| = 11$ db.

15.11. Determine the gain and phase margins for the system with the open-loop transfer function of Problem 15.8.

From Fig. 15-29, $\omega_1 = 2.3$ rad/sec. From Fig. 15-30, $\arg GH(j\omega_1) = -127°$. Hence the phase margin is $180° - 127° = 53°$. As shown in Fig. 15-30, $\arg GH(j\omega)$ approaches $-180°$ as ω decreases. Since $\arg GH(j\omega) = -180°$ only at $\omega = 0$, then $\omega_\pi = 0$. Therefore the gain margin is $-20\log_{10}|GH(j\omega_\pi)| = -12$ db using the normal procedure. Although a negative gain margin indicates instability for most systems, this system is stable, as verified by the Nyquist Stability Plot shown in Fig. 15-31. Remember that the system has an open-loop right-half plane pole; but the zero of GH at -2 acts to stabilize the system for $K = 2$.

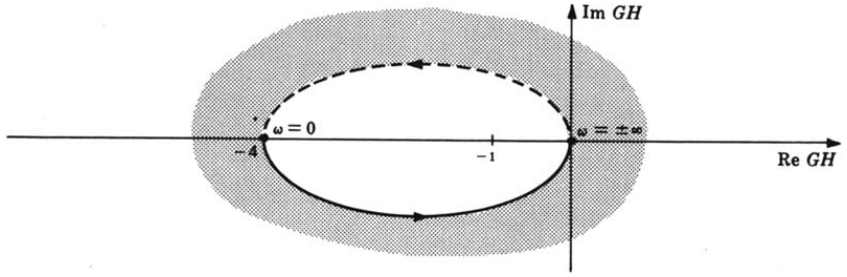


Fig. 15-31

CLOSED-LOOP FREQUENCY RESPONSE

15.12. For the system of Example 15.7 with $H = 1$, determine the closed-loop frequency response function and compare the actual closed-loop Bode magnitude plot with the approximate one of Example 15.7.

For this system, $GH = 10/s(s+1)$. Then

$$\frac{C}{R} = \frac{10}{s^2 + s + 10}$$

and

$$\frac{C}{R}(j\omega) = \frac{1}{1 + j\omega/10 - \omega^2/10}$$

Therefore the closed-loop Bode magnitude plot corresponds to Fig. 15-11, with $\zeta = 0.18$ and $\omega_n = 3.16$. From this plot the actual 3-db bandwidth is $\omega/\omega_n = 1.5$ in normalized form; hence, since $\omega_n = 3.16$, BW $= 1.5(3.16) = 4.74$ rad/sec. The approximate 3-db bandwidth determined from Fig. 15-20 of Example 15.7 is 3.7 rad/sec. Note that $\omega_n = 3.16$ rad/sec for the closed-loop system corresponds very well with $\omega_1 = 3.1$ rad/sec from Fig. 15-20. Thus the gain crossover frequency of the open-loop system corresponds very well with ω_n of the closed-loop system, although the approximate 3-db bandwidth determined above is not very accurate. The reason for this is that the approximate Bode magnitude plot of Fig. 15-20 does not show the peaking that occurs in the exact curve.

15.13. For the discrete-time system with open-loop frequency response function

$$GH(z) = \frac{3(z+1)(z+\frac{1}{3})}{8z(z-1)(z+\frac{1}{2})} \qquad H = 1$$

find the gain margin, phase margin, phase crossover angle, and gain crossover angle.

The Bode plots for this system are given in Figs. 15-32 and 15-33. The phase crossover angle $\omega_\pi T$ is determined from Fig. 15-33 as 1.74 rad. The corresponding gain margin is found on Fig. 15-32 as 11 db. The gain crossover angle $\omega_1 T$ is determined from Fig. 15-32 as 0.63 rad. The corresponding phase margin is found on Fig. 15-33 as 57°.

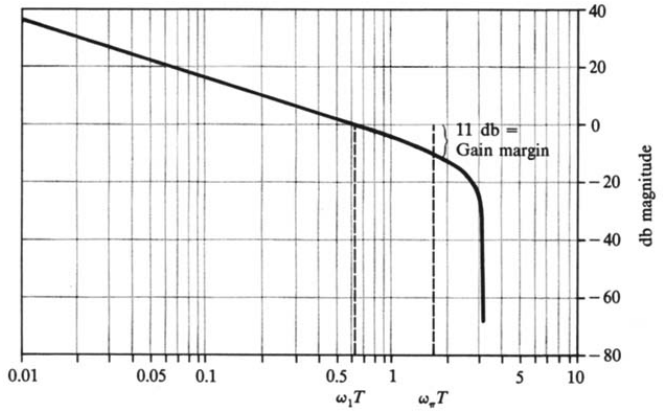


Fig. 15-32

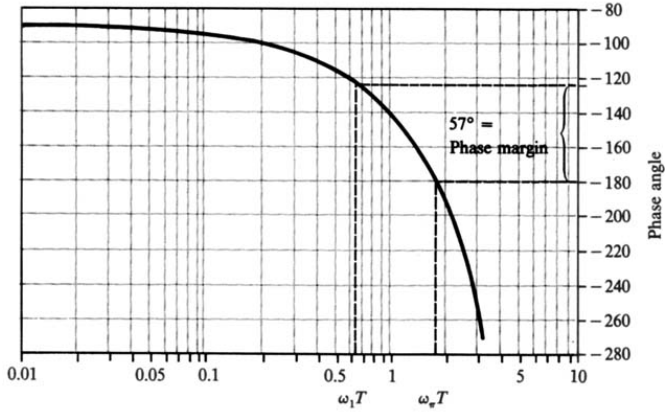


Fig. 15-33

Supplementary Problems

15.14. Construct the Bode plots for the open-loop frequency response function

$$GH(j\omega) = \frac{4(1+j\omega/2)}{(j\omega)^2(1+j\omega/8)(1+j\omega/10)}$$

15.15. Construct the Bode plots and determine the gain and phase margins for the system with the open-loop frequency response function

$$GH(j\omega) = \frac{4}{(1+j\omega)(1+j\omega/3)^2}$$

15.16. Solve Problems 13.35 and 13.37 by constructing the Bode plots.

15.17. Work Problem 13.52 using Bode plots.

15.18. Work Problem 11.59 using Bode plots.

Chapter 16

Bode Design

16.1 DESIGN PHILOSOPHY

Design of a feedback control system using Bode techniques entails shaping and reshaping the Bode magnitude and phase angle plots until the system specifications are satisfied. These specifications are most conveniently expressed in terms of frequency-domain figures of merit such as gain and phase margin for the transient performance and the error constants (Chapter 9) for the steady state time-domain response.

Shaping the asymptotic Bode plots of continuous-time systems by adding cascade or feedback compensation is a relatively simple procedure. Bode plots for several common continuous-time compensation networks are presented in Sections 16.3, 16.4, and 16.5. With these graphs, the magnitude and phase angle contributions of a particular compensator can be added directly to the uncompensated system Bode plots. It is usually necessary to correct the asymptotic Bode plots in the final stages of design to accurately verify satisfaction of the performance specifications.

Since simple asymptotic Bode plots do not exist for discrete-time systems, the shaping and reshaping of Bode plots for discrete-time systems is usually not as simple and intuitive as for continuous-time systems. However, by transforming the discrete-time transfer function into the w-plane, design of discrete-time systems can be accomplished by continuous-time techniques.

16.2 GAIN FACTOR COMPENSATION

It is possible in some cases to satisfy all system specifications by simply adjusting the open-loop gain factor K. Adjustment of the gain factor K does not affect the phase angle plot. It only shifts the magnitude plot up or down to correspond to the increase or decrease in K. The simplest procedure is to alter the db scale of the magnitude plot in accordance with the change in K instead of replotting the curve. For example, if K is doubled, the db scale should be shifted down by $20\log_{10}2 = 6.02$ db.

When working with continuous-time Bode plots, it is more convenient to use the Bode gain:

$$K_B = \frac{K\prod_{i=1}^{m} z_i}{\prod_{i=1}^{n} p_i}$$

where $-p_i$ and $-z_i$ are the finite poles and zeros of GH.

EXAMPLE 16.1. The Bode plots for

$$GH(j\omega) = \frac{K_B}{j\omega(1+j\omega/2)}$$

are shown in Fig. 16-1 for $K_B = 1$.

The maximum amount K_B may be increased to improve the system steady state performance without decreasing the phase margin below 45° is determined as follows. In Fig. 16-1, the phase margin is 45° if the gain crossover frequency ω_1 is 2 rad/sec and the magnitude plot can be raised by as much as 9 db before ω_1 becomes 2 rad/sec. Thus K_B can be increased by 9 db without decreasing the phase margin below 45°.

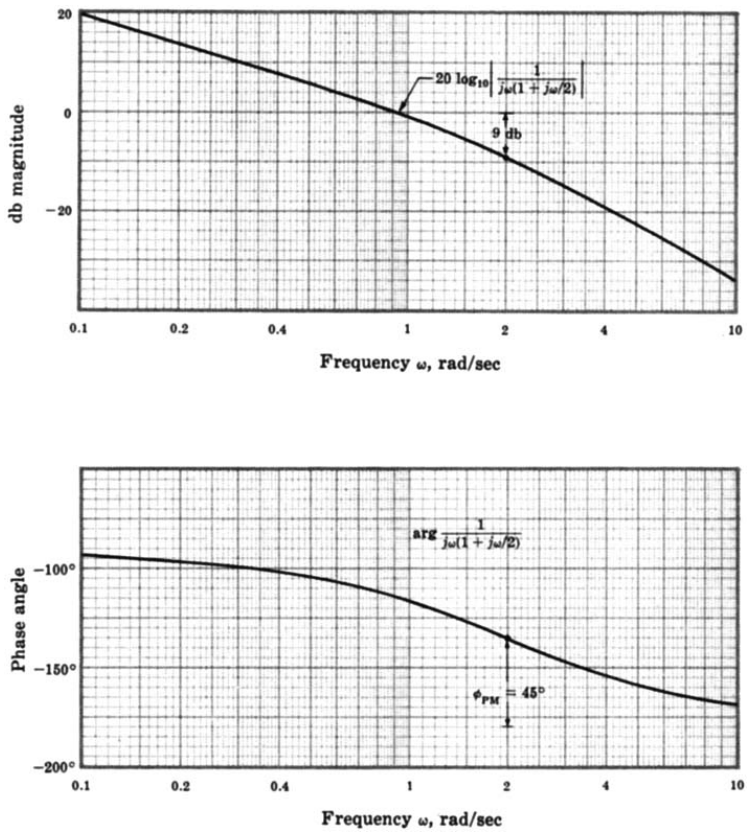


Fig. 16-1

16.3 LEAD COMPENSATION FOR CONTINUOUS-TIME SYSTEMS

The lead compensator, presented in Sections 6.3 and 12.4, has the following Bode form frequency response function:

$$P_{\text{Lead}}(j\omega) = \frac{(a/b)(1+j\omega/a)}{1+j\omega/b} \tag{16.1}$$

The Bode plots for this compensator, for various *lead ratios* a/b, are shown in Fig. 16-2. These graphs illustrate that addition of a cascade lead compensator to a system lowers the overall magnitude curve in the low-frequency region and raises the overall phase angle curve in the low-to-mid-frequency region. Other properties of the lead compensator are discussed in Section 12.4.

The amount of low-frequency attenuation and phase lead produced by the lead compensator depends on the lead ratio a/b. Maximum phase lead occurs at the frequency $\omega_m = \sqrt{ab}$ and is equal to

$$\phi_{\max} = \left(90 - 2\tan^{-1}\sqrt{a/b}\right) \text{ degrees} \tag{16.2}$$

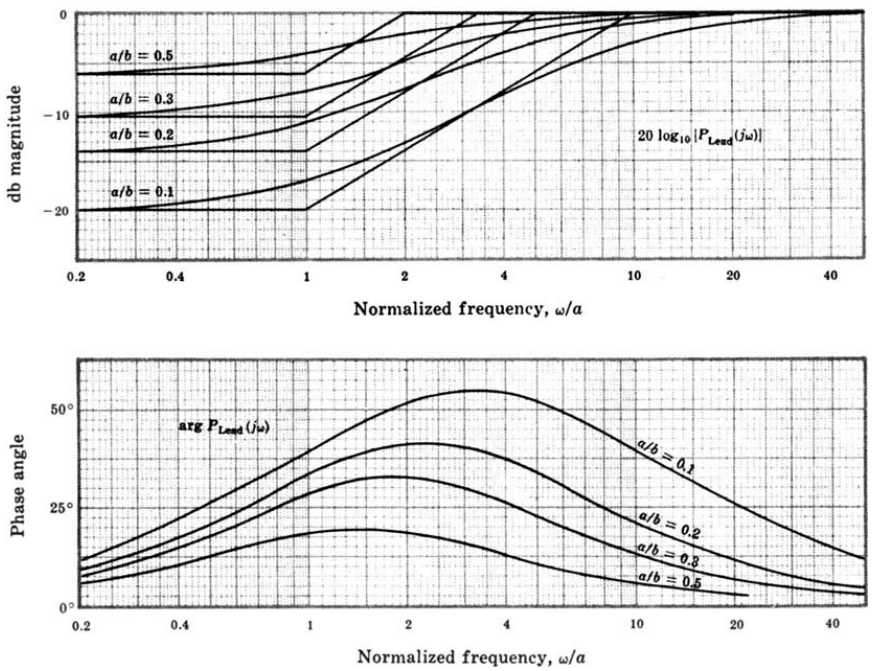


Fig. 16-2

Lead compensation is normally used to increase the gain and/or phase margins of a system or increase its bandwidth. An additional modification of the Bode gain K_B is usually required with lead networks, as described in Section 12.4.

EXAMPLE 16.2. An uncompensated continuous-time system whose open-loop transfer function is

$$GH = \frac{24}{s(s+2)(s+6)} \qquad H = 1$$

is to be designed to meet the following performance specifications:

1. when the input is a ramp with slope (velocity) 2π rad/sec, the steady state position error must be less than or equal to $\pi/10$ radians.
2. $\phi_{\text{PM}} = 45° \pm 5°$.
3. gain crossover frequency $\omega_1 \geq 1$ rad/sec.*

Lead compensation is appropriate, as previously outlined in detail in Example 12.4. Transforming $GH(j\omega)$ into Bode form,

$$GH(j\omega) = \frac{2}{j\omega(1+j\omega/2)(1+j\omega/6)}$$

we note that the Bode gain K_B is equal to the velocity error constant $K_{v1} = 2$. The Bode plots for this system are shown in Fig. 16-3.

*When using Bode techniques, closed-loop system *bandwidth* specifications are often interpreted in terms of the gain crossover frequency ω_1, which is easily determined from the Bode magnitude plot. The bandwidth and ω_1 are not generally equivalent; but, when one increases or decreases, the other usually follows. As noted in Sections 10.4, and 15.8 and Problem 12.16, ω_1 is often a reasonable approximation for the bandwidth.

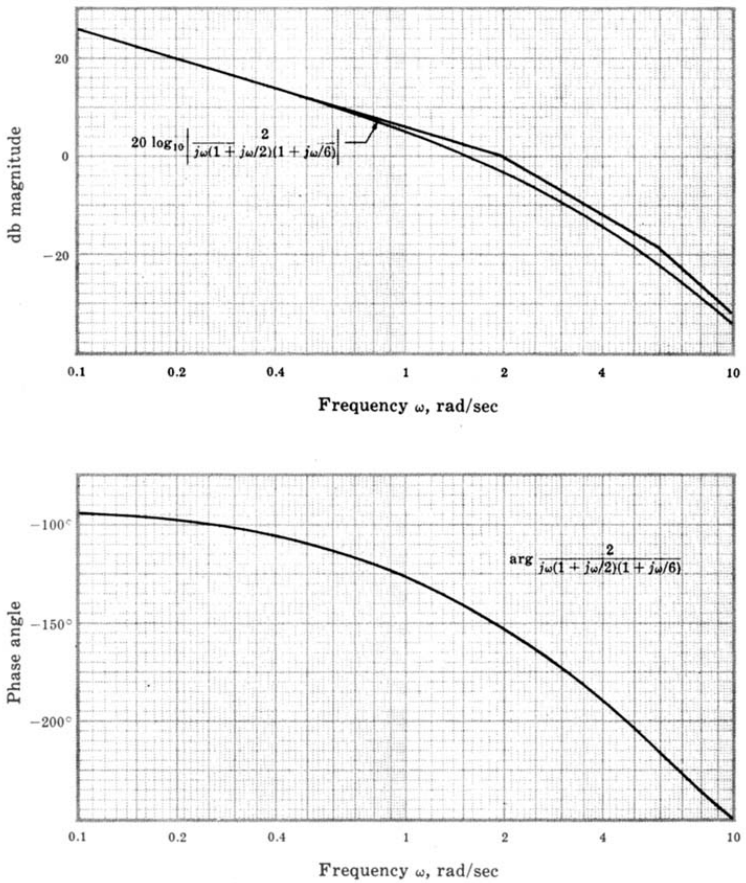


Fig. 16-3

The steady state error $e(\infty)$ is given by Equation (*9.13*) as $1/K_v$ for a unit ramp function input. Therefore, if $e(\infty) \leq \pi/10$ radians and the ramp has a slope of 2π instead of 1, then the required velocity error constant is

$$K_{v2} \geq \frac{2\pi}{\pi/10} = 20 \text{ sec}^{-1}$$

Thus a cascade amplifier with a gain of $\lambda = 10$, or 20 db, will satisfy the steady state specification. But this gain must be further increased after the lead network parameters are chosen, as described in Example 12.4. When the Bode gain is increased by 20 db, the gain margin is -8 db and the phase margin $-28°$, as read directly from the plots of Fig. 16-3. Therefore the lead compensator must be chosen to bring the phase margin to 45°. This requires a large amount of phase lead. Furthermore, since addition of the lead compensator must be accompanied by an increase in gain of b/a, the net effect is to increase the gain at mid and high frequencies, thus raising the gain crossover frequency. Hence a phase margin of 45° has to be established at a higher frequency, requiring even more phase lead. For these reasons we add two cascaded lead networks (with the necessary isolation to reduce loading effects, if required).

To determine the parameters of the lead compensator, assume that the Bode gain has been increased by 20 db so that the 0-db line is effectively lowered by 20 db. If we choose $b/a = 10$, then the lead compensator plus an

additional Bode gain increase of $(b/a)^2$ for the two networks has the following combined form:

$$[10P_{\text{Lead}}(j\omega)]^2 = G_c(j\omega) = \frac{(1 + j\omega/a)^2}{(1 + j\omega/10a)^2}$$

Now we must choose an appropriate value for a. A useful method for improving system stability is to try to cross the 0-db line at a slope of -6 db/octave. Crossing at a slope of -12 db/octave usually results in too low a value for the phase margin. If a is set equal to 2, a sketch of the asymptotes reveals that the 0-db line is crossed at -12 db/octave. If $a = 4$, the 0-db line is crossed at a slope of -6 db/octave. The Bode magnitude and phase angle plots for the system with $a = 4$ rad/sec are shown in Fig. 16-4. The gain margin is 14 db and the phase margin is 50°. Thus the second specification is satisfied. The gain crossover frequency $\omega_1 = 14$ rad/sec is substantially higher than the value specified, indicating that the system will respond a good deal faster than required by the third specification. The compensated system block diagram is shown in Fig. 16-5. A properly designed amplifier may additionally serve the purpose of load-effect isolation if it is placed *between* the two lead networks.

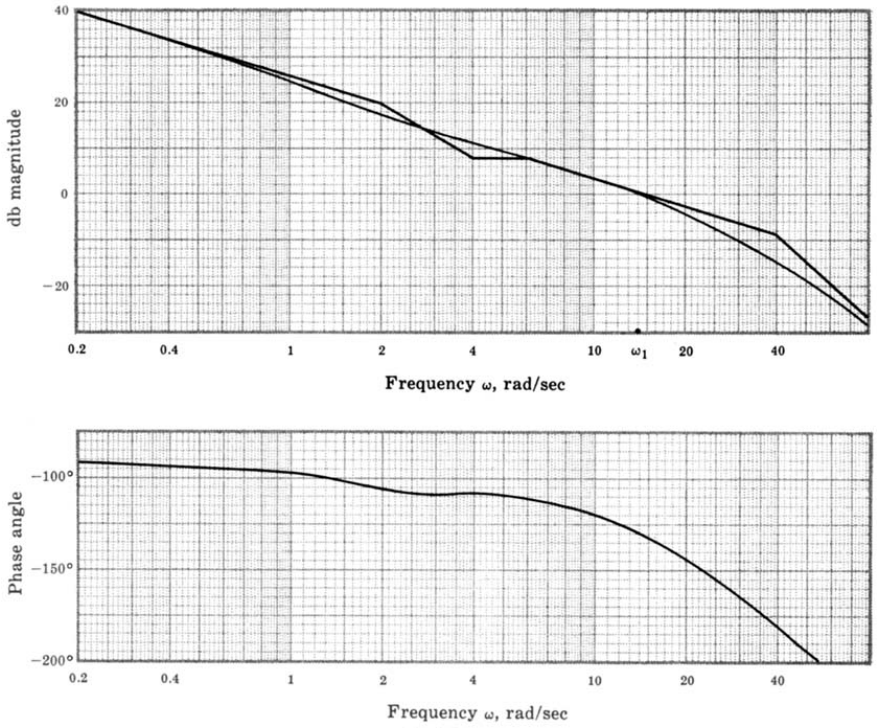


Fig. 16-4

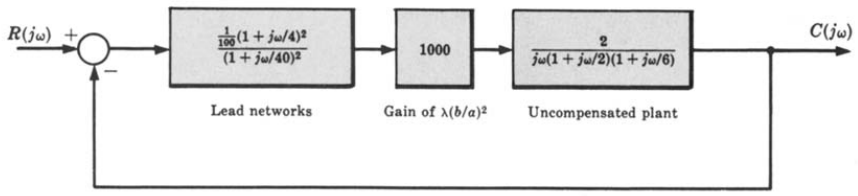


Fig. 16-5

16.4 LAG COMPENSATION FOR CONTINUOUS-TIME SYSTEMS

The lag compensator, presented in Sections 6.3 and 12.5, has the following Bode form frequency response function:

$$P_{\text{Lag}}(j\omega) = \frac{1 + j\omega/b}{1 + j\omega/a} \tag{16.3}$$

The Bode plots for the lag compensator, for several *lag ratios* b/a, are shown in Figure 16-6. The properties of this compensator are discussed in Section 12.5.

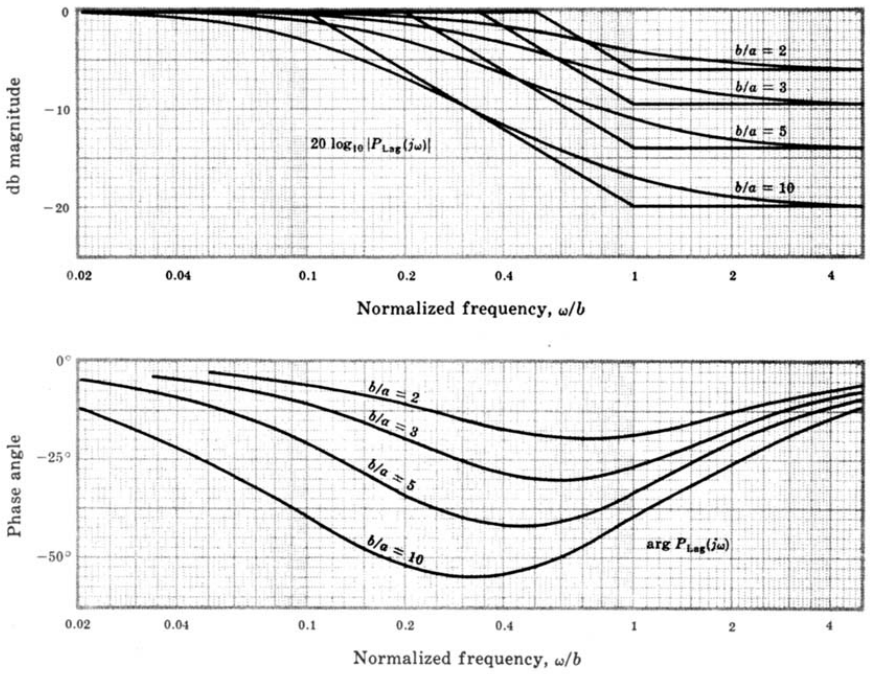


Fig. 16-6

EXAMPLE 16.3. Let us redesign the system of Example 16.2 using gain factor plus *lag* compensation, as previously outlined in detail in Example 12.5. The uncompensated system is, again, represented by

$$GH(j\omega) = \frac{2}{j\omega(1 + j\omega/2)(1 + j\omega/6)}$$

and the specifications are

1. $K_v \geq 20\ \text{sec}^{-1}$
2. $\phi_{\text{PM}} = 45° \pm 5°$
3. $\omega_1 \geq 1$ rad/sec

As before, a Bode gain increase by a factor of 10, or 20 db, is required to satisfy the first (steady state) specification. Hence the Bode plots of Fig. 16-3 should again be considered with the 0-db line effectively lowered by 20 db. Addition of significant phase-lag at frequencies less than 0.1 rad/sec will lower the curve or effectively raise the 0-db line by an amount corresponding to b/a. Thus the ratio b/a must be chosen so that the resulting phase

margin is 45°. From the Bode phase angle plot (Fig. 16-3) we see that a 45° phase margin is obtained if the gain crossover frequency is $\omega_1 = 1.3$ rad/sec. From the Bode magnitude plot, this requires that the magnitude curve be lowered by 2 + 20 = 22 db. Thus a gain decrease of 22 db, or a factor of 14, is needed. This can be obtained using a lag compensator with $b/a = 14$. The actual location of the compensator is arbitrary as long as the phase shift produced at ω_1 is negligible. Values of $a = 0.01$ and $b = 0.14$ rad/sec are adequate. The compensated system block diagram is shown in Fig. 16-7.

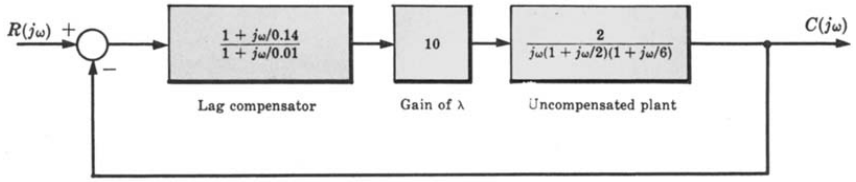


Fig. 16-7

16.5 LAG-LEAD COMPENSATION FOR CONTINUOUS-TIME SYSTEMS

It is sometimes desirable, as discussed in Section 12.6, to simultaneously employ both lead and lag compensation. Although one each of these two networks can be connected in series to achieve the desired effect, it is usually more convenient to mechanize the combined lag-lead compensator described in Example 6.6. This compensator can be constructed with a single *R-C* network, as shown in Problem 6.14.

The Bode form of the frequency response function for the lag-lead compensator is

$$P_{\text{LL}}(j\omega) = \frac{(1 + j\omega/a_1)(1 + j\omega/b_2)}{(1 + j\omega/b_1)(1 + j\omega/a_2)}$$

with $b_1 > a_1$, $b_2 > a_2$ and $a_1 b_2 = b_1 a_2$. A typical Bode magnitude plot in which $a_1 > b_2$ is shown in Fig. 16-8. The Bode plots for a specific lag-lead compensator can be determined by combining the Bode plots for the lag portion from Fig. 16-6 with those for the lead portion from Fig. 16-2. Additional properties of the lag-lead compensator are discussed in Section 12.6.

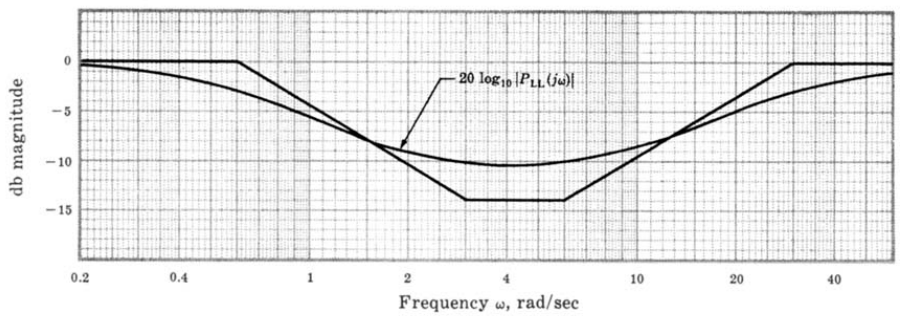


Fig. 16-8

EXAMPLE 16.4. Let us redesign the system of Example 16.2 using lag-lead compensation. Suppose, for example, that we want the gain crossover frequency ω_1 (approximate closed-loop bandwidth) to be greater than 2 rad/sec but less than 5 rad/sec, with all the other specifications the same as Example 16.2. For this application, we shall see that the lag-lead compensator has advantages over either lag or lead compensation. The uncompensated system is, again, represented by

$$GH(j\omega) = \frac{2}{j\omega(1 + j\omega/2)(1 + j\omega/6)}$$

The Bode plots are shown in Fig. 16-3. As in Example 16.2, a Bode gain increase of 20 db is required to satisfy the specification on steady state performance. Once again referring to Fig. 16-3 with the 0-db line shifted down by 20 db to correspond to the Bode gain increase, the parameters of the lag-lead compensator must be chosen to result in a gain crossover frequency between 2 and 5 rad/sec, with a phase margin of about 45°. The phase angle plot of Fig. 16-3 shows about $-188°$ phase angle at approximately 4 rad/sec. Thus we need about 53° phase lead to establish a 45° phase margin in that frequency range. Let us choose a lead ratio of $a_1/b_1 = 0.1$ to make sure we have enough phase lead. To place it in about the right frequency range, let $a_1 = 0.8$ and $b_1 = 8$ rad/sec. The lag portion must have the same ratio $a_2/b_2 = 0.1$; but the lag portion must be sufficiently lower than a_1 so as not to significantly reduce the phase lead obtained from the lead portion; $b_2 = 0.2$ and $a_2 = 0.02$ are adequate. The Bode plots for the compensated system are shown in Fig. 16-9; and the block diagram is shown in Fig. 16-10.

We note that the lag-lead compensator produces no magnitude attenuation at either high or low frequencies. Therefore a smaller gain factor adjustment (as obtained with lag compensation in Example 16.3) and a smaller bandwidth and gain crossover frequency (as that resulting from lead compensation in Example 16.2) are obtained using lag-lead compensation.

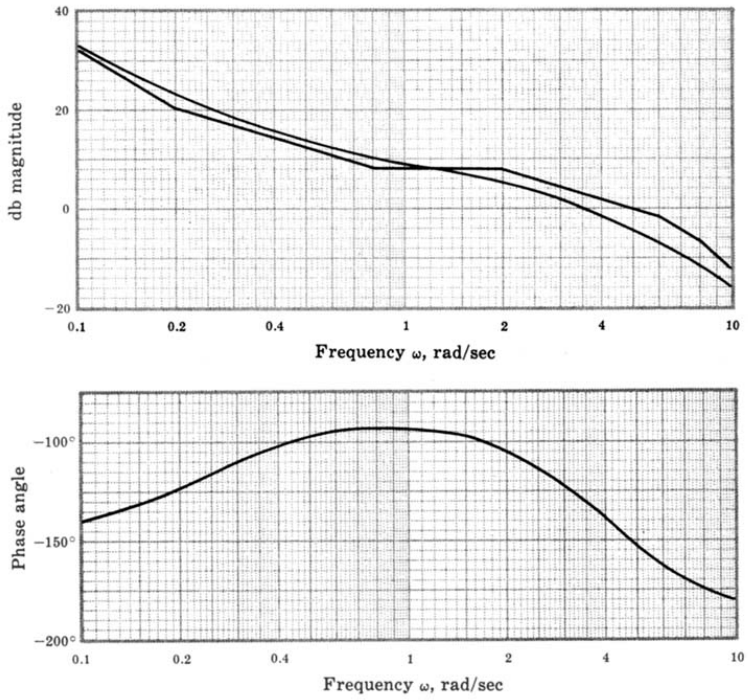


Fig. 16-9

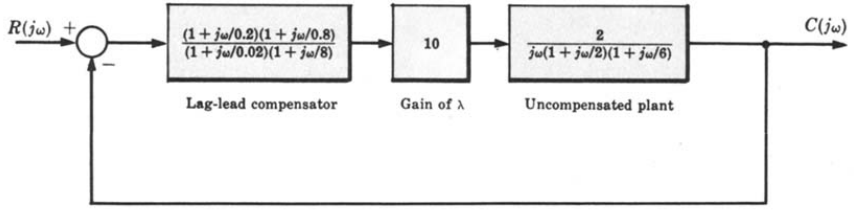


Fig. 16-10

16.6 BODE DESIGN OF DISCRETE-TIME SYSTEMS

Bode design of discrete-time systems is based on the same philosophy as Bode design of continuous-time systems in that it entails shaping and reshaping the Bode magnitude and phase angle plots until the system specifications are met. But the effort involved can be substantially greater.

It is sometimes possible to satisfy specifications by simply adjusting the open-loop gain factor K, as described in Section 16.2 for continuous-time systems.

EXAMPLE 16.5. Consider the discrete-time system of Example 15.4, with open-loop frequency response function

$$GH(e^{j\omega T}) = \frac{\frac{1}{100}(e^{j\omega T}+1)^2}{(e^{j\omega T}-1)(e^{j\omega T}+\frac{1}{3})(e^{j\omega T}+\frac{1}{2})}$$

and $H = 1$. Figures 16-11 and 16-12 are the Bode plots of GH, drawn by computer, which illustrate the gain and

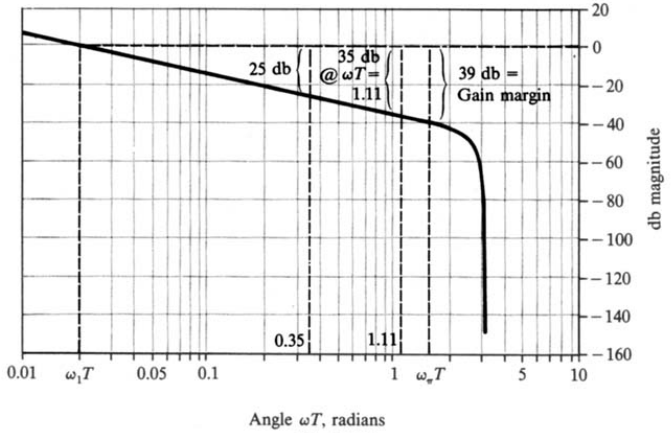


Fig. 16-11

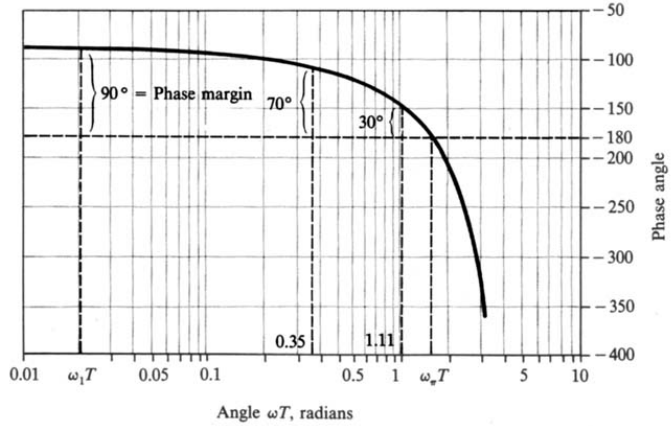


Fig. 16-12

phase margins and the gain and phase crossover frequencies. We now show that gain factor compensation alone can be used to satisfy the following system specifications:

1. $\phi_{PM} \geq 30°$.
2. 10 db $\leq$ gain margin $\leq$ 15 db.

From Fig. 16-12 we see if $\omega_1 T$ can be increased to 1.11 rad, then $\phi_{PM} = 30°$. To accomplish this, the gain must be increased by 35 db, as shown in Fig. 16-11, resulting in a gain margin of $39 - 35 = 4$ db, which is too small. If we increase the gain by only 25 db (increase K by a factor of 18), then $\omega_1 T = 0.35$ rad and the phase margin is 70°. Note that changing K does not change $\omega_n T$.

For discrete-time system design specifications which cannot be satisfied by gain factor compensation alone, Bode design in the z-domain is not as straightforward as in the s-domain. Continuous-time system design methods can, however, be transferred to the design of discrete-time systems using the *w-transform*. Based on developments in Sections 10.7 and 15.9, the design algorithm is as follows:

1. Substitute $(1 + w)/(1 - w)$ for z in the open-loop transfer function $GH(z)$:
$$GH(z)|_{z=(1+w)/(1-w)} \equiv GH'(w)$$
2. Set $w = j\omega_w$, and then transform critical frequencies in the performance specifications from the z- to the w-domain, using:
$$\omega_w = \tan\frac{\omega T}{2}$$
3. Develop continuous-time compensation (as in Sections 16.3 through 16.5) such that the system in the w-domain satisfies the given specifications at the frequencies obtained in Step 2 (as if the w-domain were the s-domain).
4. Transform the compensation elements obtained in Step 3 back to the z-domain to complete the design, using $w = (z - 1)/(z + 1)$.

EXAMPLE 16.6. The unity feedback discrete-time system with open-loop transfer function

$$G(z) = GH(z) = \frac{3}{8}\frac{(z+1)(z+\frac{1}{3})}{z(z+\frac{1}{2})}$$

and sampling period $T = 0.1$ sec is to be compensated so that it meets the following specifications:

1. The steady state error must be less than or equal to 0.02 for a unit ramp input.
2. $\phi_{PM} \geq 30°$.
3. The gain crossover frequency ω_1 must satisfy $\omega_1 T \geq 1$ rad.

This is a type 0 system and the steady state error for a unit ramp input is infinite (Section 9.9). Therefore the compensation must contain a pole at $z = 1$ and the new transfer function including this pole becomes

$$GH'(z) = \frac{3}{8}\frac{(z+1)(z+\frac{1}{3})}{z(z-1)(z+\frac{1}{2})}$$

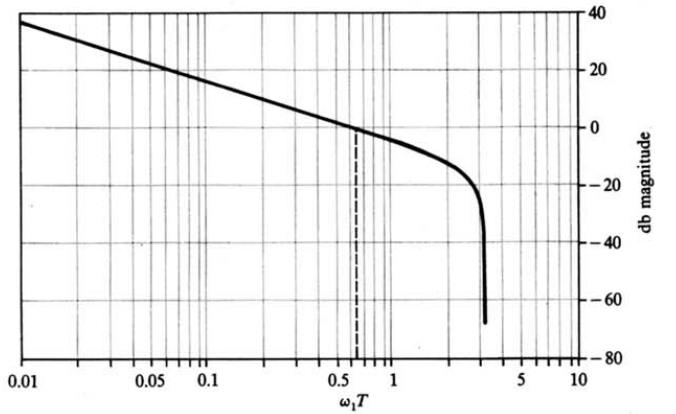
From the table in Section 9.9 the steady state error for the unit ramp is $e(\infty) = T/K_v$, where $K_v = GH(1) = \lim_{z\to 1}(z-1)GH'(z) = \frac{2}{3}$. Thus, with $e(\infty) = 0.15$, the gain factor must be increased by a factor of 15/2 (17.5 db).

The Bode plots for GH' are shown in Figs. 16-13 and 16-14. From Fig. 16-13, the angle at the gain crossover frequency is $\omega_1 T = 0.68$ rad and the phase margin is 56°. Increasing the gain by 17.5 db would move the angle at the gain crossover frequency to $\omega_1 T = 2.56$ rad, but the phase margin would then become $-41°$, destabilizing the system. Gain factor compensation alone is apparently inadequate for this design problem.

To complete the design, we transform $GH(z)$ into the w-domain, setting $z = (1 + w)/(1 - w)$ and forming

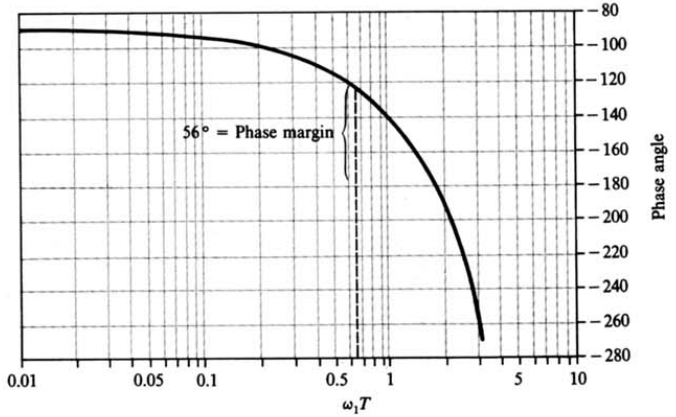
$$GH''(w) = \frac{1}{3}\frac{(1-w)(1+w/2)}{w(1+w)(1+w/3)}$$

The Bode plots for GH'' are shown in Figs. 16-15 and 16-16.



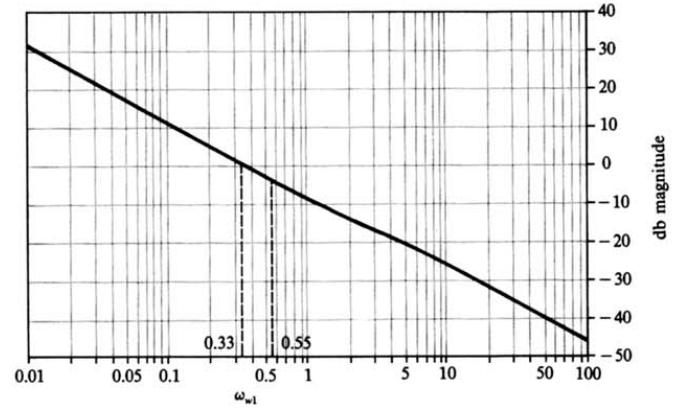
Angle ωT, radians

Fig. 16-13



Angle ωT, radians

Fig. 16-14



Frequency ω_w, rad/sec

Fig. 16-15

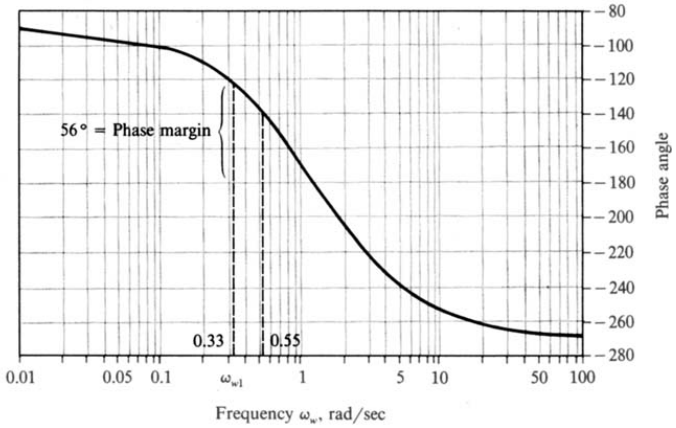


Fig. 16-16

Following Step 2 above, the gain crossover frequency specification $\omega_1 T \geq 1$ rad is transformed into the w-plane using

$$\omega_{w1} = \tan\frac{\omega_1 T}{2} \geq \tan\frac{1}{2} = 0.55 \text{ rad/sec}$$

From Fig. 16-15 [or from $\omega_{w1} = \tan(0.68/2)$] the gain crossover frequency is 0.35 rad/sec and the phase margin is 56° (as it was in the z-domain).

To satisfy the steady state error specification, the gain factor must be increased by at least 17.5 db (as noted earlier), and to satisfy the remaining specifications, the gain crossover frequency should be increased to at least 0.55 rad/sec (Fig. 16-16), and the phase angle at $\omega_w = 0.55$ should be held to at least $-150°$. This last requirement implies that no more than 6.5° of lag can be introduced at $\omega_w = 0.55$ rad/sec. Note that this requires about 4.3-db gain increase at $\omega_w = 0.55$ rad/sec so that this frequency can become the gain crossover frequency.

Lag compensation can satisfy these specifications (Step 3). From Fig. 16-6, a lag ratio of $b/a = 5$ provides 14 db of attenuation at higher frequencies. To increase the gain crossover frequency, the gain factor is increased by 18.3 db, so that at $\omega_w = 0.55$ there is a net increase of 4.3 db. This is clearly adequate to also satisfy the steady state error specification (17.5 db is needed).

Now the parameter a in the lag ratio can be chosen to satisfy the phase margin requirement. As noted above, we must keep the phase lag of the compensator below 6.5° at $\omega_w = 0.55$ rad/sec. We note that the phase lag of the lag compensator is

$$\phi_{\text{Lag}} = \tan^{-1}\frac{\omega T}{b} - \tan^{-1}\frac{\omega T}{a}$$

Thus, setting $\phi_{\text{Lag}} = -6.5°$, $\omega = \omega_w = 0.55$ rad/sec and $b = 5a$ (as above), this equation is easily solved for a. Choosing the smaller of the solutions generates a *dipole* (a pole-zero pair) very near the origin of the w-plane, for $a = 0.0157$. We choose $a = 0.015$ which gives only 6.2° of phase lag. Thus $b = 0.075$ and the lag compensator in the w-plane is given by

$$P_{\text{Lag}}(w) = \left(\frac{0.015}{0.075}\right)\left(\frac{w + 0.075}{w + 0.015}\right)$$

P_{Lag} is now transformed back to the z-domain (Step 4) by substituting $w = (z-1)/(z+1)$. The result is

$$P_{\text{Lag}}(z) = 0.21182\left(\frac{z - 0.86046}{z - 0.97044}\right)$$

Combining this with the pole at $z = 1$ and the gain factor increase of 18.3 db (a gain factor ratio increase of 8.22),

the complete compensation element $G_1(z)$ is

$$G_1(z) = 1.7417\left[\frac{z - 0.86046}{(z-1)(z - 0.97044)}\right]$$

The compensated control system is shown in Fig. 16-17. Note that this design is quite similar to those developed for this same system and specifications in Examples 12.7 and 14.5.

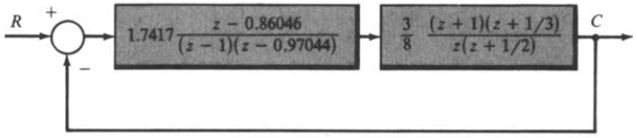


Fig. 16-17

Solved Problems

GAIN FACTOR COMPENSATION

16.1. Determine the maximum value for the Bode gain K_B which will result in a gain margin of 6 db or more and a phase margin of 45° or more for the system with the open-loop frequency response function

$$GH(j\omega) = \frac{K_B}{j\omega(1 + j\omega/5)^2}$$

The Bode plots for this system with $K_B = 1$ are shown in Fig. 16-18.

The gain margin, measured at $\omega_\pi = 5$ rad/sec, is 20 db. Thus the Bode gain can be raised by as much as $20 - 6 = 14$ db and still satisfy the gain margin requirement. However, the Bode phase angle plot indicates that, for $\phi_{\text{PM}} \geq 45°$, the gain crossover frequency ω_1 must be less than about 2 rad/sec. The magnitude curve can be raised by as much as 7.5 db before ω_1 exceeds 2 rad/sec. Thus the maximum value of K_B satisfying *both* specifications is 7.5 db, or 2.37.

16.2. Design the system of Problem 15.7, to have a phase margin of 55°.

The Bode phase angle plot in Fig. 15-28 indicates that the gain crossover frequency ω_1 must be 0.9 rad/sec for 55° phase margin. From the Bode magnitude of Fig. 15-27, K_B must be reduced by 6 db, or a factor of 2, to achieve $\omega_1 = 0.9$ rad/sec and hence $\phi_{\text{PM}} = 55°$.

LEAD COMPENSATION

16.3. Show that the maximum phase lead of the lead compensator [Equation (*16.1*)] occurs at $\omega_m = \sqrt{ab}$ and prove Equation (*16.2*).

The phase angle of the lead compensator is $\phi = \arg P_{\text{Lead}}(j\omega) = \tan^{-1}\omega/a - \tan^{-1}\omega/b$. Then

$$\frac{d\phi}{d\omega} = \frac{1}{a\left[1 + (\omega/a)^2\right]} - \frac{1}{b\left[1 + (\omega/b)^2\right]}$$

Setting $d\phi/d\omega$ equal to zero yields $\omega^2 = ab$. Thus the maximum phase lead occurs at $\omega_m = \sqrt{ab}$. Hence $\phi_{\max} = \tan^{-1}\sqrt{b/a} - \tan^{-1}\sqrt{a/b}$. But since $\tan^{-1}\sqrt{b/a} = \pi/2 - \tan^{-1}\sqrt{a/b}$, we have $\phi_{\max} = (90 - 2\tan^{-1}\sqrt{a/b})$ degrees.

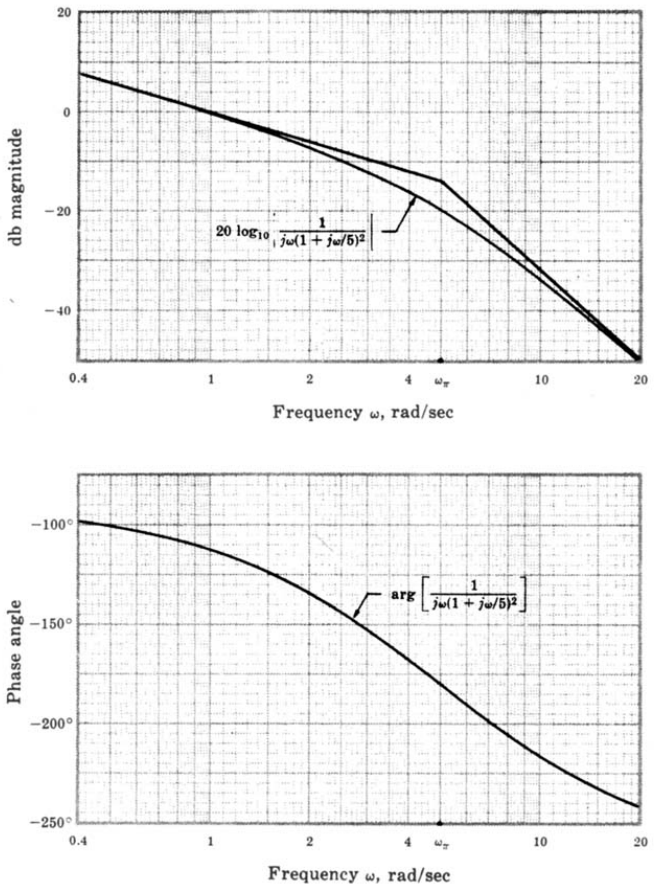


Fig. 16-18

16.4. What attenuation (magnitude) is produced by a lead compensator at the frequency of maximum phase lead $\omega_m = \sqrt{ab}$?

The attenuation factor is given by

$$\left| P_{\text{Lead}}(j\sqrt{ab}) \right| = \left| \frac{(a/b)(1+j\sqrt{b/a})}{(1+j\sqrt{a/b})} \right| = \frac{a}{b}\sqrt{\frac{1+b/a}{1+a/b}} = \sqrt{\frac{a}{b}}$$

16.5. Design compensation for the system

$$GH(j\omega) = \frac{8}{(1+j\omega)(1+j\omega/3)^2}$$

which will yield an overall phase margin of 45° and the same gain crossover frequency ω_1 as the uncompensated system. The latter is essentially the same as designing for the same bandwidth, as discussed in Section 15.8.

The Bode plots for the uncompensated system are shown in Fig. 16-19.

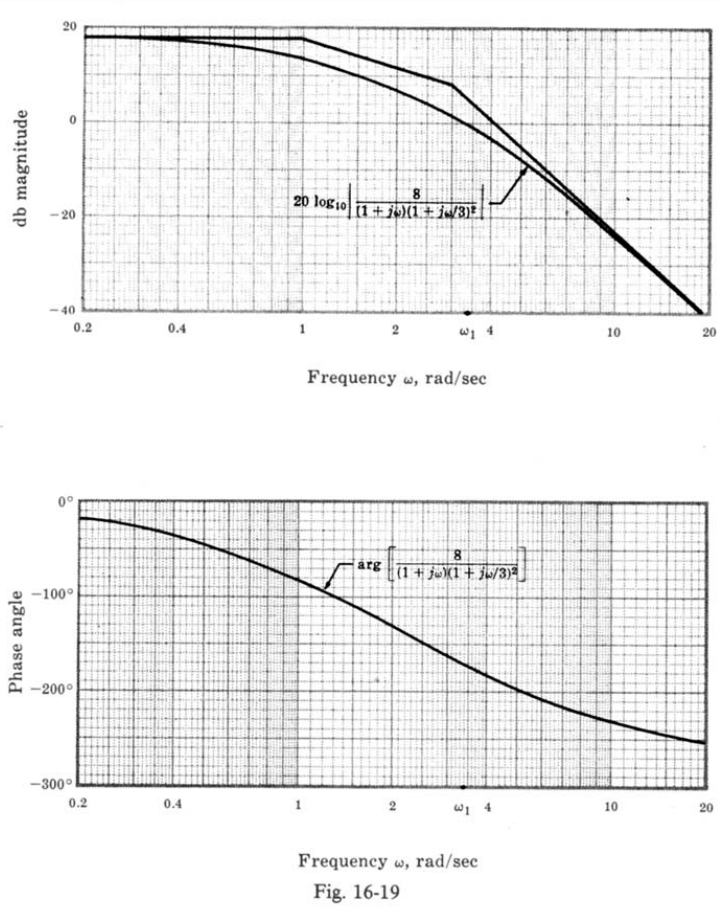


Fig. 16-19

The gain crossover frequency ω_1 is 3.4 rad/sec and the phase margin is 10°. The specifications can be met with a cascade lead compensator and gain factor amplifier. Choosing a and b for the lead compensator is somewhat arbitrary, as long as the phase lead at $\omega_1 = 3.4$ is sufficient to raise the phase margin from 10° to 45°. However, it is often desirable, for economic reasons, to minimize the low-frequency attenuation obtained from the lead network by choosing the largest lead ratio $a/b < 1$ that will supply the required amount of phase lead. Assuming this is the case, the maximum lead ratio that will yield 45° − 10° = 35° phase lead is about 0.3 from Fig. 16-2. Solution of Equation (*16.2*) yields a value of $a/b = 0.27$. But we shall use $a/b = 0.3$ because we have the curves available for this value in Fig. 16-2. We want to choose a and b such that the maximum phase lead, which occurs at $\omega_m = \sqrt{ab}$, is obtained at $\omega_1 = 3.4$ rad/sec. Thus $\sqrt{ab} = 3.4$. Substituting $a = 0.3b$ into this equation and solving for b, we find $b = 6.2$ and $a = 1.86$. But this compensator produces $20\log_{10}\sqrt{6.2/1.86} = 5.2$ db attenuation at $\omega_1 = 3.4$ rad/sec (see Problem 16.4). Thus an amplifier with a gain of 5.2 db, or 1.82, is required, in addition to the lead compensator, to maintain ω_1 at 3.4 rad/sec. The Bode plots for the compensated system are shown in Fig. 16-20 and the block diagram in Fig. 16-21.

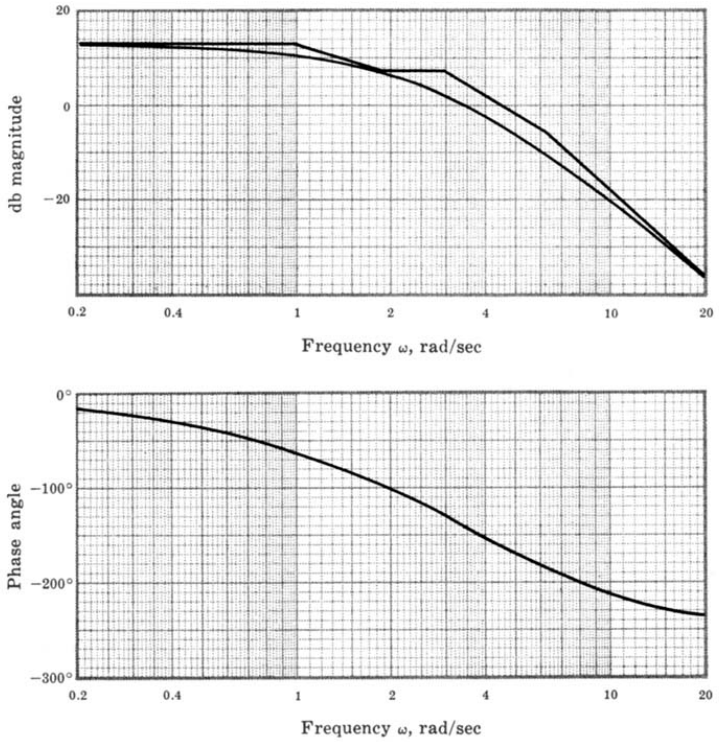


Fig. 16-20

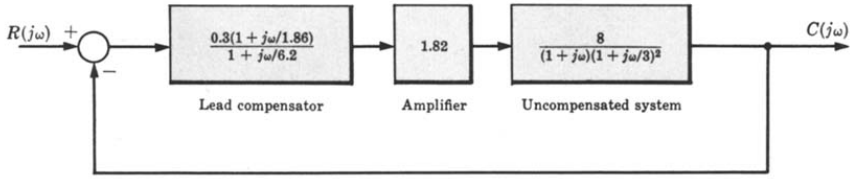


Fig. 16-21

LAG COMPENSATION

16.6. What is the maximum phase lag produced by the lag compensator [Equation (*16.3*)]?

The phase angle of the lag compensator is

$$\arg P_{\text{Lag}}(j\omega) = \tan^{-1}\frac{\omega}{b} - \tan^{-1}\frac{\omega}{a} = -\arg P_{\text{Lead}}(j\omega)$$

Thus the maximum phase lag (negative phase angle) of the lag compensator is the same as the maximum phase lead of the lead compensator with the same values of a and b. Hence the maximum also occurs at $\omega_m = \sqrt{ab}$ and, from Equation (*16.2*), we get

$$\phi_{\max} = \left(90 - 2\tan^{-1}\sqrt{\frac{a}{b}}\right) \text{ degrees}$$

Expressed in terms of the lag ratio b/a, this equation becomes

$$\phi_{\max} = \left(2\tan^{-1}\sqrt{\frac{b}{a}} - 90\right) \text{ degrees}$$

16.7. Design compensation for the system of Problem 16.1 to satisfy the same specifications and, in addition, to have a gain crossover frequency ω_1 less than or equal to 1 rad/sec and a velocity error constant $K_v > 5$.

The Bode plots for this system, shown in Fig. 16-18, indicate that $\omega_1 = 1$ rad/sec for $K_B = 1$. Hence $K_v = K_B = 1$ for $\omega_1 = 1$. The gain and phase margin requirements are easily met with any $K_B < 2.37$; but the steady state specification requires $K_v = K_B > 5$. Therefore a low-frequency cascade lag compensator with $b/a = 5$ can be used to increase K_v to 5, while maintaining the crossover frequency and the gain and phase margins at their previous values. A lag compensator with $b = 0.5$ and $a = 0.1$ satisfies this requirements, as shown in Fig. 16-22.

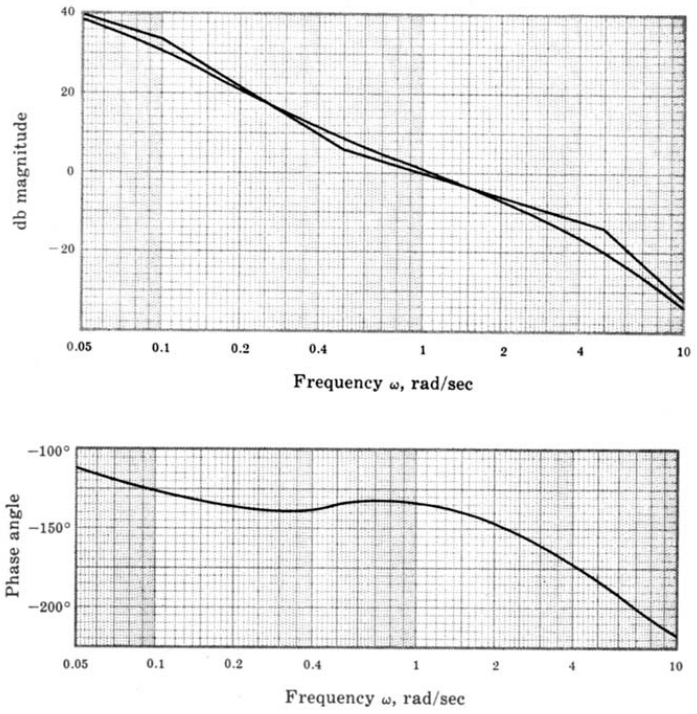


Fig. 16-22

The compensated open-loop frequency response function is $\dfrac{5(1+j\omega/0.5)}{j\omega(1+j\omega/0.1)(1+j\omega/5)^2}$.

16.8. Design a discrete-time unity feedback system, with the fixed plant

$$G_2(z) = \frac{27}{64}\frac{(z+1)^3}{\left(z+\frac{1}{2}\right)^3}$$

satisfying the specifications: (1) $K_p \geq 4$, (2) gain margin ≥ 12 db, (3) phase margin $\geq 45°$.

The specification on the position error constant K_p requires a gain factor increase of 4. This transfer function is transformed into the w-plane by letting $z = (1 + w)/(1 - w)$ thus forming

$$G_2'(w) = \frac{1}{(1 + w/3)^3}$$

The Bode plots for this system, with the gain factor increased by $20\log_{10}4 = 12$ db, are shown in Fig. 16-23.

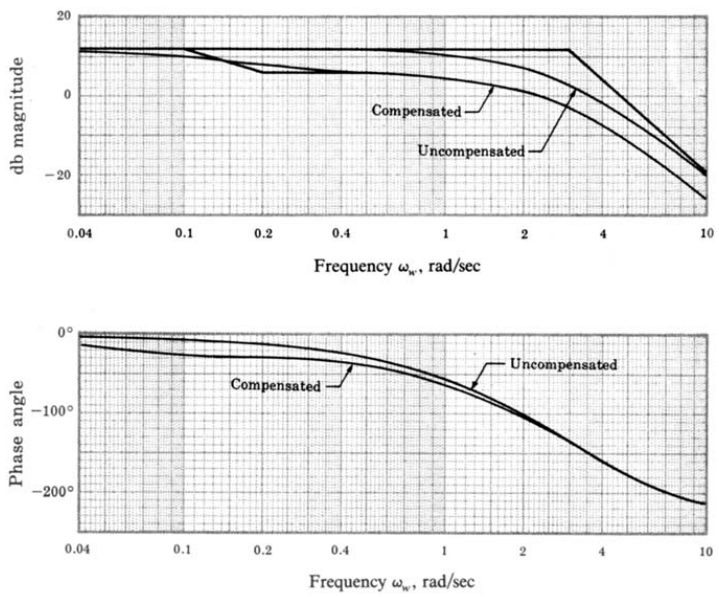


Fig. 16-23

The gain margin is 6 db and the phase margin is 30°. These margins can be increased by adding a lag compensator. To increase the gain margin by 12 db, the high-frequency magnitude must be reduced by 6 db. To raise the phase margin to 45°, ω_{w1} must be lowered to 3.0 rad/sec or less. This requires a magnitude attenuation of 3 db at that frequency. Therefore let us choose a lag ratio $b/a = 2$ to yield a high-frequency attenuation of $20\log_{10}2 = 6$ db. For $a = 0.1$ and $b = 0.2$ the phase margin is 65° and the gain margin is 12 db, as shown in the compensated Bode plots of Fig. 16-23.

The compensated open-loop frequency response function is

$$\frac{4(1 + j\omega_w/0.2)}{(1 + j\omega_w/0.1)(1 + j\omega_w)^3}$$

The compensation element

$$G_1'(w) = \frac{4(1 + w/0.2)}{1 + w/0.1}$$

is transformed back to the z-domain by letting $w = (z - 1)/(z + 1)$ thus forming

$$G_1(z) = \frac{24}{11}\frac{\left(z - \frac{2}{3}\right)}{\left(z - \frac{9}{11}\right)}$$

LAG-LEAD COMPENSATION

16.9. Determine compensation for the system of Problem 16.5 that will result in a position error constant $K_p \geq 10$, $\phi_{PM} \geq 45°$ and the same gain crossover frequency ω_1 as the uncompensated system.

The compensation determined in Problem 16.5 satisfies all the specifications except that K_p is only 4.4 The lead compensator chosen in that problem has a low-frequency attenuation of 10.4 db, or a factor of 3.33. Let us replace the lead network with a lag-lead compensator, choosing $a_1 = 1.86$, $b_1 = 6.2$, and $a_2/b_2 = 0.3$. The low-frequency magnitude becomes $a_1b_2/b_1a_2 = 1$, or 0 db, and the attenuation produced by the lead network is erased, effectively raising K_p for the system by a factor of 3.33 to 14.5. The lag portion of the compensator should be placed at frequencies sufficiently low so that the phase margin is not reduced below the specified value of 45°. This can be accomplished with $a_2 = 0.09$ and $b_2 = 0.3$. The compensated system block diagram is shown in Fig. 16-24. Note that an amplifier with a gain of 1.82 is included, as in Problem 16.5, to maintain $\omega_1 = 3.4$.

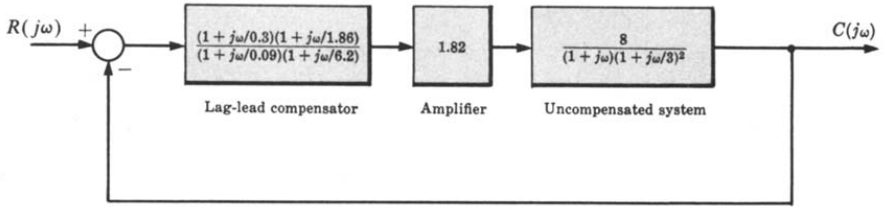


Fig. 16-24

The compensated Bode plots are shown in Fig. 16-25.

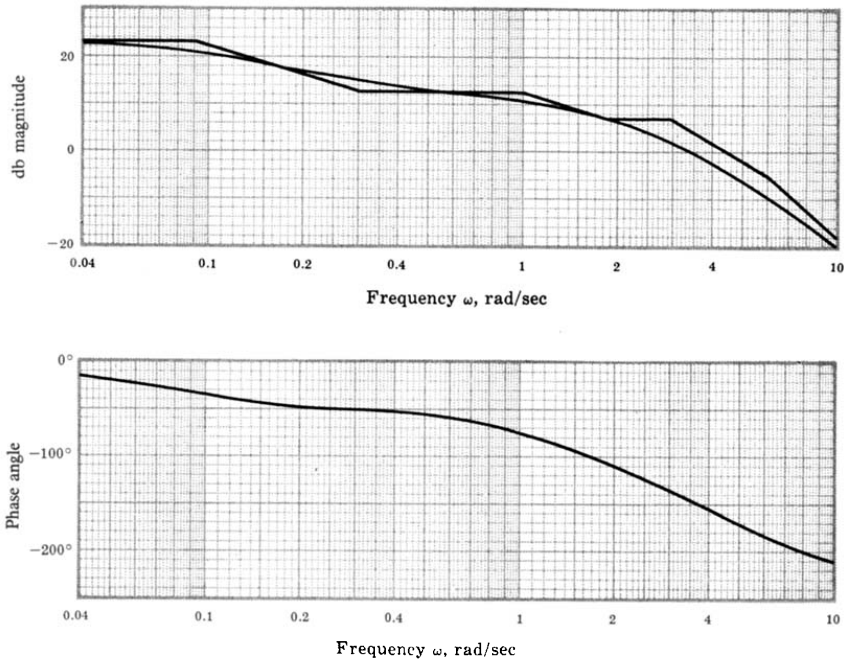


Fig. 16-25

16.10. Design cascade compensation for a unity feedback control system, with the plant

$$G_2(j\omega) = \frac{1}{j\omega(1 + j\omega/8)(1 + j\omega/20)}$$

to meet the following specifications:

(1) $K_v \geq 100$ (3) gain margin ≥ 10 db

(2) $\omega_1 \geq 10$ rad/sec (4) phase margin $\phi_{PM} \geq 45°$

To satisfy the first specification, a Bode gain increase by a factor of 100 is required since the uncompensated $K_v = 1$. The Bode plots for this system, with the gain increased to 100, are shown in Fig. 16-26.

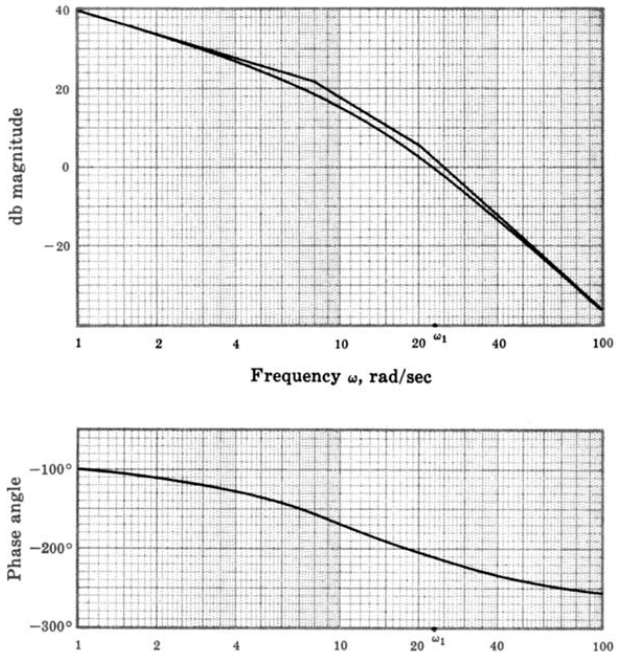


Fig. 16-26

The gain crossover frequency $\omega_1 = 23$ rad/sec, the phase margin is $-30°$, and the gain margin is -12 db. Lag compensation could be used to increase the gain and phase margins by reducing ω_1. However, ω_1 would have to be lowered to less than 8 rad/sec to achieve a 45° phase margin and to less than 6 rad/sec for a 10-db gain margin. Consequently, we would not satisfy the second specification. With lead compensation, an additional Bode gain increase by a factor of b/a would be required and ω_1 would be increased, thus requiring substantially more than the 75° phase lead for $\omega_1 = 23$ rad/sec. These disadvantages can be overcome using lag-lead compensation. The lead portion produces attenuation and phase lead. The frequencies at which these effects occur must be positioned near ω_1 so that ω_1 is slightly reduced and the phase margin is increased. Note that, although *pure* lead compensation increases ω_1, the lead portion of lag-lead compensator decreases ω_1 because the gain factor increase of b/a is unnecessary, thereby lowering the magnitude characteristic. The lead portion can be determined independently using the curves of Fig. 16-2; but it must be kept in mind that, when the lag portion is included, the attenuation and phase lead

may be somewhat reduced. Let us try a lead ratio of $a_1/b_1 = 0.1$, with $a_1 = 5$ and $b_1 = 50$. The maximum phase lead then occurs at 15.8 rad/sec. This enables the magnitude asymptote to cross the 0-db line with a slope of -6 db/octave (see Example 16.2). The compensated Bode plots are shown in Fig. 16-27 with a_2 and b_2 chosen as 0.1 and 1.0 rad/sec, respectively. The resulting parameters are $\omega_1 = 12$ rad/sec, gain margin = 14 db, and $\phi_{PM} = 52°$, as shown on the graphs. The compensated open-loop frequency response function is

$$\frac{100(1 + j\omega)(1 + j\omega/5)}{j\omega(1 + j\omega/0.1)(1 + j\omega/8)(1 + j\omega/20)(1 + j\omega/50)}$$

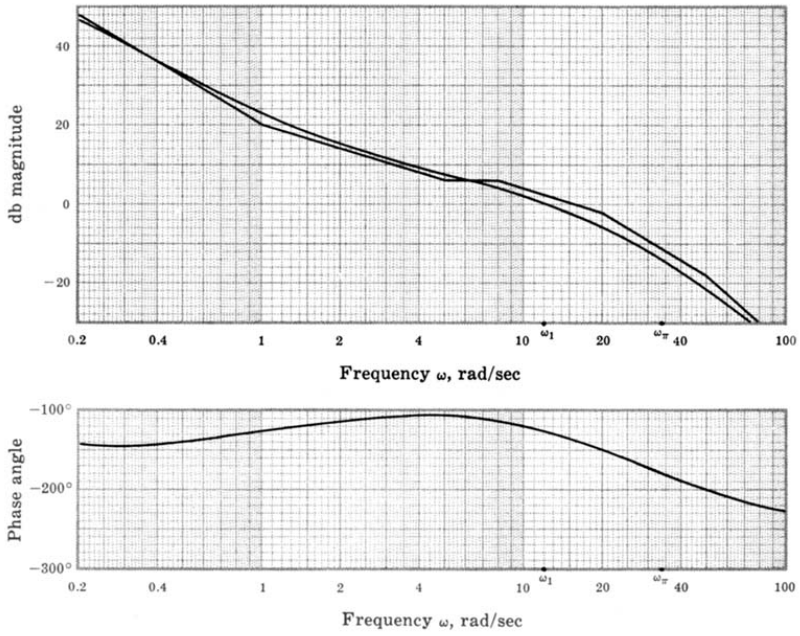


Fig. 16-27

MISCELLANEOUS PROBLEM

16.11. The nominal frequency response function of a certain plant is

$$G_2(j\omega) = \frac{1}{j\omega(1 + j\omega/8)(1 + j\omega/20)}$$

A feedback control system must be designed to control the output of this plant for a certain application and it must satisfy the following frequency domain specifications:

(*1*) gain margin ≥ 6 db

(*2*) phase margin $(\phi_{PM}) \geq 30°$

In addition, it is known that the "fixed" parameters of the plant may vary slightly during operation of the system. The effects of this variation on the system response must be minimized over the frequency range of interest, which is $0 \leq \omega \leq 8$ rad/sec, and the actual requirement can

be interpreted as a specification on the sensitivity of $(C/R)(j\omega)$ with respect to $|G_2(j\omega)|$, that is,

$$(3) \qquad 20\left|\log_{10} S^{(C/R)(j\omega)}_{|G_2(j\omega)|}\right| \le -10 \text{ db} \qquad \text{for} \quad 0 \le \omega \le 8 \text{ rad/sec}$$

It is also known that the plant will be subjected to an uncontrollable, additive disturbance input, represented in the frequency domain by $U(j\omega)$. For this application, the system response to this disturbance input must be suppressed in the frequency range $0 \le \omega \le 8$ rad/sec. Therefore the design problem includes the additional constraint on the magnitude ratio of the output to the disturbance input given by

$$(4) \qquad 20\log_{10}\left|\frac{C}{U}(j\omega)\right| \le -20 \text{ db} \qquad \text{for} \quad 0 \le \omega \le 8 \text{ rad/sec}$$

Design a system which satisfies these four specifications.

The general system configuration, which includes the possibility of either or both cascade and feedback compensators, is shown in Fig. 16-28.

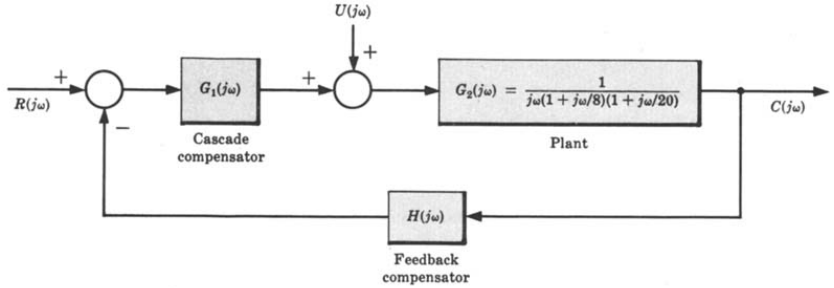


Fig. 16-28

From Fig. 16-28, we see that

$$\frac{C}{U}(j\omega) = \frac{G_2(j\omega)}{1 + G_1G_2H(j\omega)} \qquad \text{and} \qquad \frac{C}{R}(j\omega) = \frac{G_1G_2(j\omega)}{1 + G_1G_2H(j\omega)}$$

In a manner similar to that of Example 9.7, it is easily shown that

$$S^{(C/R)(j\omega)}_{|G_2(j\omega)|} = \frac{1}{1 + G_1G_2H(j\omega)}$$

If we assume that $|G_1G_2H(j\omega)| \gg 1$ in the frequency range $0 \le \omega \le 8$ rad/sec (this inequality must be checked upon completion of the design and, if it is not satisfied, the compensation may have to be recomputed), then specification (3) may be approximated by

$$20\log_{10}\left|S^{(C/R)(j\omega)}_{|G_2(j\omega)|}\right| \cong 20\log_{10}\left|\frac{1}{G_1G_2H(j\omega)}\right|$$

$$= -20\log_{10}|G_1G_2H(j\omega)| \le -10 \text{ db}$$

or

$$20\log_{10}|G_1G_2H(j\omega)| \ge 10 \text{ db}$$

Similarly, specification (4) can be approximated by

$$20\log_{10}\left|\frac{C}{U}(j\omega)\right| \cong 20\log_{10}\frac{|G_2(j\omega)|}{|G_1G_2H(j\omega)|}$$

$$= 20\log_{10}|G_2(j\omega)| - 20\log_{10}|G_1G_2H(j\omega)| \le -20 \text{ db}$$

or

$$20\log_{10}|G_1G_2H(j\omega)| \ge \left[20 + 20\log_{10}|G_2(j\omega)|\right] \text{ db}$$

Specifications (3) and (4) can therefore be translated into the following combined form. We require that the open-loop frequency response, $G_1G_2H(j\omega)$, lie in a region on a Bode magnitude plot which simultaneously satisfies the two inequalities:

$$20\log_{10}|G_1G_2H(j\omega)| \ge 10 \text{ db}$$

$$20\log_{10}|G_1G_2H(j\omega)| \ge \left[20 + 20\log_{10}|G_2(j\omega)|\right] \text{ db}$$

This region lies above the broken line shown in the Bode magnitude plot in Fig. 16-29, which also includes Bode plots of $G_2(j\omega)$. The design may be completed by determining compensation which satisfies the gain and phase margin requirements, (1) and (2), subject to this magnitude constraint.

A 32-db increase in Bode gain, which is necessary at $\omega = 8$ rad/sec, would satisfy specifications (3) and (4), but not (1) and (2). Therefore a more complicated compensation is required. For a second trial, we find that the lag-lead compensation:

$$G_1H'(j\omega) = \frac{100(1 + j\omega/2.5)(1 + j\omega/0.25)}{(1 + j\omega/25)(1 + j\omega/0.025)}$$

results in a system with a gain margin of 6 db and $\phi_{PM} \cong 26°$, as shown in Fig. 16-29. We see from the figure that 10° to 15° more phase lead is necessary near $\omega = 25$ rad/sec and $|G_1H'(j\omega)|$ must be increased by at least 2 db in the neighborhood of $\omega = 8$ rad/sec to satisfy the magnitude constraint. If we introduce an additional lead network and increase the Bode gain to compensate for the low-frequency attenuation of the lead network, the compensation becomes

$$G_1H''(j\omega) = 300\left(\frac{1 + j\omega/10}{1 + j\omega/30}\right)\left[\frac{(1 + j\omega/2.5)(1 + j\omega/0.25)}{(1 + j\omega/25)(1 + j\omega/0.025)}\right]$$

This results in a gain margin of 7 db, $\phi_{PM} \cong 30°$, and satisfaction of specifications (3) and (4), as shown in Fig. 16-29. The assumption that $|G_1G_2H(j\omega)| \gg 1$ for $0 \le \omega \le 8$ rad/sec is easily shown to be justified by

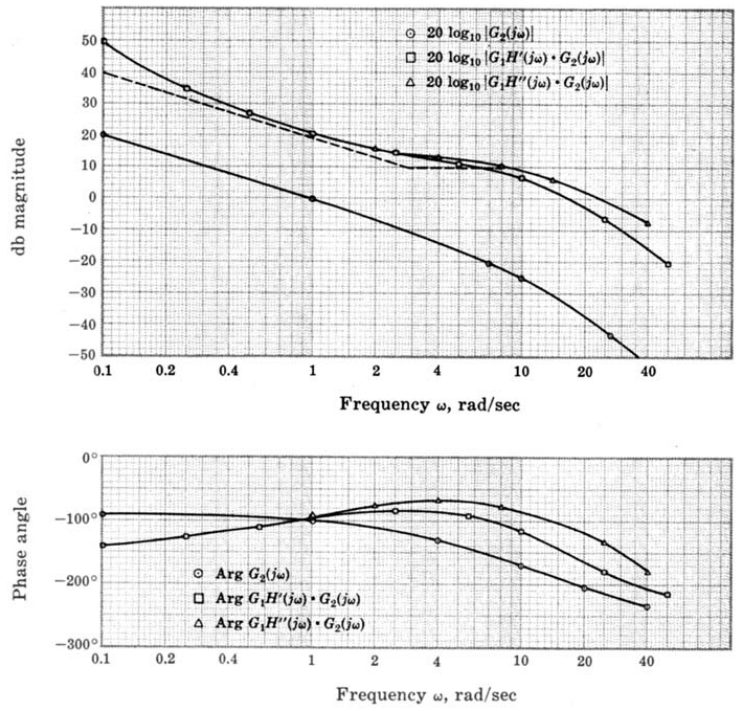


Fig. 16-29

calculating the actual values of the db magnitudes of

$$\left|S^{(C/R)(j\omega)}_{|G_2(j\omega)|}\right| \quad \text{and} \quad \left|\frac{C}{U}(j\omega)\right|$$

The compensator $G_1H''(j\omega)$ can be divided between the forward and feedback paths, or put all in one path, depending on the form desired for $(C/R)(j\omega)$ if such a form is specified by the application.

Supplementary Problems

16.12. Design a compensator for the system with the open-loop frequency response function

$$GH(j\omega) = \frac{20}{j\omega(1+j\omega/10)(1+j\omega/25)(1+j\omega/40)}$$

to result in a closed-loop system with a gain margin of at least 10 db and a phase margin of at least 45°.

16.13. Determine a compensator for the system of Problem 16.1 which will result in the same gain and phase margins but with a crossover frequency ω_1 of at least 4 rad/sec.

16.14. Design a compensator for the system with the open-loop frequency response function

$$GH(j\omega) = \frac{2}{(1+j\omega)\left[1+j\omega/10-(\omega/4)^2\right]}$$

which will result in a closed-loop system with a gain margin of at least 6 db and a phase margin of at least 40°.

16.15. Work Problem 12.9 using Bode plots. Assume a maximum of 25% overshoot will be ensured if the system has a phase margin of at least 45°.

16.16. Work Problem 12.10 using Bode plots.

16.17. Work Problem 12.20 using Bode plots.

16.18. Work Problem 12.21 using Bode plots.

Chapter 17

Nichols Chart Analysis

17.1 INTRODUCTION

Nichols chart analysis, a frequency response method, is a modification of the Nyquist and Bode methods. The *Nichols chart* is essentially a transformation of the M- and N-circles on the Polar Plot (Section 11.12) into noncircular M and N contours on a db magnitude versus phase angle plot in rectangular coordinates. If $GH(\omega)$ represents the open-loop frequency response function of either a continuous-time or discrete-time system, then $GH(\omega)$ plotted on a Nichols chart is called a *Nichols chart plot* of $GH(\omega)$. The relative stability of the closed-loop system is easily obtained from this graph. The determination of absolute stability, however, is generally impractical with this method and either the techniques of Chapter 5 or the Nyquist Stability Criterion (Section 11.10) are preferred.

The reasons for using Nichols chart analysis are the same as those for the other frequency response methods, the Nyquist and Bode techniques, and are discussed in Chapters 11 and 15. The Nichols chart plot has at least two advantages over the Polar Plot: (1) a much wider range of magnitudes can be graphed because $|GH(\omega)|$ is plotted on a logarithmic scale; and (2) the graph of $GH(\omega)$ is obtained by algebraic summation of the individual magnitude and phase angle contributions of its poles and zeros. While both of these properties are also shared by Bode plots, $|GH(\omega)|$ and $\arg GH(\omega)$ are included on a single Nichols chart plot rather than on two Bode plots.

Nichols chart techniques are useful for directly plotting $(C/R)(\omega)$ and are especially applicable in system design, as shown in the next chapter.

17.2 db MAGNITUDE-PHASE ANGLE PLOTS

The polar form of both continuous-time and discrete-time open-loop frequency response functions is

$$GH(\omega) = |GH(\omega)| \angle \arg GH(\omega) \qquad (17.1)$$

Definition 17.1: The **db magnitude-phase angle plot** of $GH(\omega)$ is a graph of $|GH(\omega)|$, in decibels, versus $\arg GH(\omega)$, in degrees, on rectangular coordinates with ω as a parameter.

EXAMPLE 17.1. The db magnitude-phase angle plot of the continuous-time open-loop frequency response function

$$GH(j\omega) = 1 + j\omega = \sqrt{1+\omega^2} \angle \tan^{-1}\omega$$

is shown in Fig. 17-1.

17.3 CONSTRUCTION OF db MAGNITUDE-PHASE ANGLE PLOTS

The db magnitude-phase angle plot for either a continuous-time or discrete-time system can be constructed directly by evaluating $20\log_{10}|GH(\omega)|$ and $\arg GH(\omega)$ in degrees, for a sufficient number of values of ω (or ωT) and plotting the results in rectangular coordinates with the log magnitude as the ordinate and the phase angle as the abscissa. Available software makes this a relatively simple process.

EXAMPLE 17.2. The db magnitude-phase angle plot of the open-loop frequency response function

$$GH(e^{j\omega T}) = \frac{\frac{1}{100}(e^{j\omega T}+1)^2}{(e^{j\omega T}-1)\left(e^{j\omega T}+\frac{1}{3}\right)\left(e^{j\omega T}+\frac{1}{2}\right)}$$

is shown in Fig. 17-2. Note that ωT is the parameter along the curve.

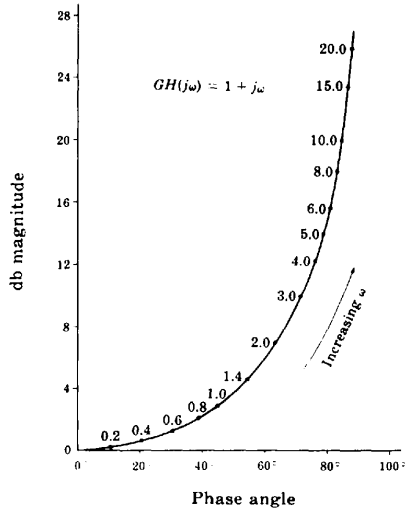


Fig. 17-1

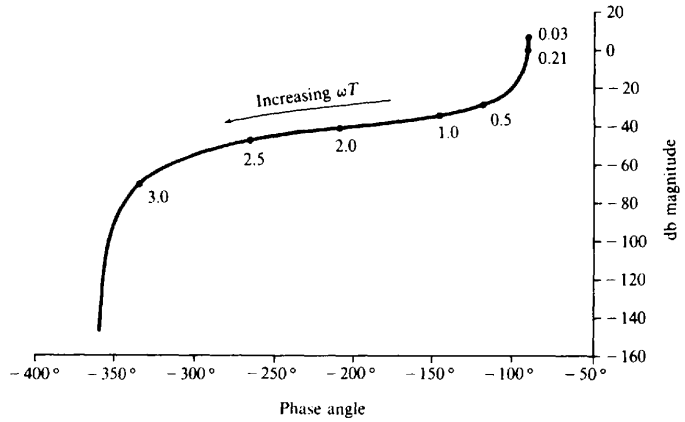


Fig. 17-2

A graphical approach to construction of db magnitude-phase angle plots is illustrated by examining the technique for continuous-time systems.

First write $GH(j\omega)$ in the *Bode form* (Section 15.3):

$$GH(j\omega) = \frac{K_B(1+j\omega/z_1)\cdots(1+j\omega/z_m)}{(j\omega)^l(1+j\omega/p_1)\cdots(1+j\omega/p_n)}$$

where l is a nonnegative integer. For $K_B > 0$ [if $K_B < 0$, add $-180°$ to $\arg GH(j\omega)$],

$$20\log_{10}|GH(j\omega)| = 20\log_{10}K_B + 20\log_{10}\left|1+\frac{j\omega}{z_1}\right| + \cdots + 20\log_{10}\left|1+\frac{j\omega}{z_m}\right| + 20\log_{10}\left|\frac{1}{(j\omega)^l}\right| + 20\log_{10}\left|\frac{1}{1+\frac{j\omega}{p_1}}\right| + \cdots + 20\log_{10}\left|\frac{1}{1+\frac{j\omega}{p_n}}\right| \quad (17.2)$$

$$\arg GH(j\omega) = \arg\left(1+\frac{j\omega}{z_1}\right) + \cdots + \arg\left(1+\frac{j\omega}{z_m}\right) + \arg\left[\frac{1}{(j\omega)^l}\right] + \arg\frac{1}{1+\frac{j\omega}{p_1}} + \cdots + \arg\frac{1}{1+\frac{j\omega}{p_n}} \quad (17.3)$$

Using Equations (*17.2*) and (*17.3*), the db magnitude-phase angle plot of $GH(j\omega)$ is generated by summing the db magnitudes and phase angles of the poles and zeros, or pairs of poles and zeros when they are complex conjugates.

The db magnitude-phase angle plot of K_B is a straight line parallel to the phase angle axis. The ordinate of the straight line is $20\log_{10}K_B$.

The db magnitude-phase angle plot for a *pole of order l at the origin*,

$$\frac{1}{(j\omega)^l} \quad (17.4)$$

is a straight line parallel to the db magnitude axis with an abscissa $-90l°$ as shown in Fig. 17-3. Note that the parameter along the curve is ω^l.

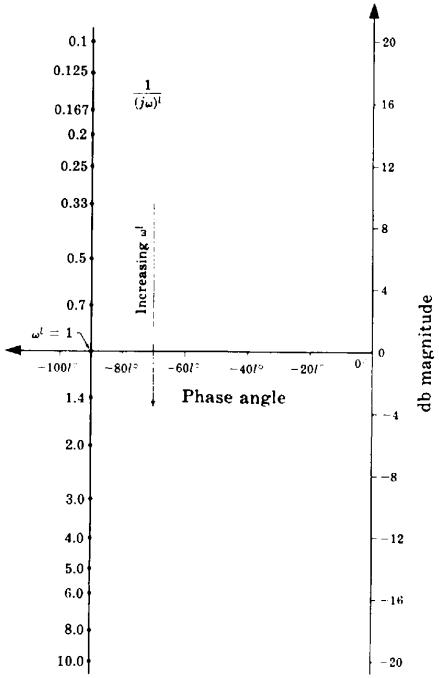


Fig. 17-3

The plot for a *zero of order l at the origin*,

$$(j\omega)^l \tag{17.5}$$

is a straight line parallel to the db magnitude axis with an abscissa of $90l°$. The plot for $(j\omega)^l$ is the diagonal mirror image about the origin of the plot for $1/(j\omega)^l$. That is, for fixed ω the db magnitude and phase angle of $1/(j\omega)^l$ are the negatives of those for $(j\omega)^l$.

The db magnitude-phase angle plot for a *real pole*,

$$\frac{1}{1+j\omega/p} \qquad p>0 \tag{17.6}$$

is shown in Fig. 17-4. The shape of the graph is independent of p because the frequency parameter along the curve is normalized to ω/p.

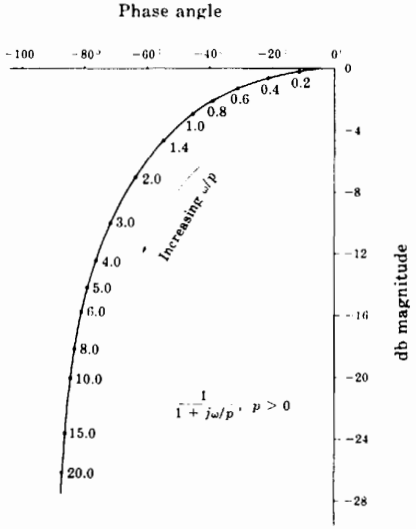


Fig. 17-4

The plot for a *real zero*,

$$1+\frac{j\omega}{z} \qquad z>0 \tag{17.7}$$

is the diagonal mirror image about the origin of Fig. 17-4.

A set of db magnitude-phase angle plots of several pairs of *complex conjugate poles*,

$$\frac{1}{1-(\omega/\omega_n)^2+j2\zeta(\omega/\omega_n)} \qquad 0<\zeta<1 \tag{17.8}$$

are shown in Fig. 17-5. For fixed ζ, the graphs are independent of ω_n because the frequency parameter is normalized to ω/ω_n.

The plots for *complex conjugate zeros*,

$$1-\left(\frac{\omega}{\omega_n}\right)^2+j2\zeta\left(\frac{\omega}{\omega_n}\right) \qquad 0<\zeta<1 \tag{17.9}$$

are diagonal mirror images about the origin of Fig. 17-5.

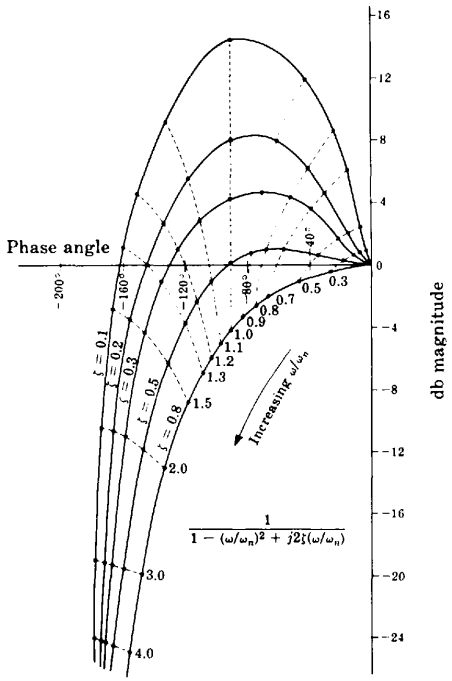


Fig. 17-5

EXAMPLE 17.3. The db magnitude-phase angle plot of

$$GH(j\omega)=\frac{10(1+j\omega/2)}{(1+j\omega)\left[1-(\omega/2)^2+j\omega/2\right]}$$

is constructed by adding the db magnitudes and phase angles of the individual factors:

$$10 \qquad 1+\frac{j\omega}{2} \qquad \frac{1}{1+j\omega} \qquad \frac{1}{1-(\omega/2)^2+j\omega/2}$$

Tabulation of these factors is helpful, as in Table 17.1. The first row contains the db magnitude and phase

Table 17.1

Term \ Frequency ω	0	0.4	0.8	1.2	1.6	2	2.8	4	6	8
10	20 db 0°	20 0°	20 0°	20 0°	20 0°	20 0°	20 0°	20 0°	20 0°	20 0°
$1+\frac{j\omega}{2}$	0 db 0°	0.2 11°	0.6 21°	1.3 31°	2.2 39°	3.0 45°	4.7 54°	7 63°	10 71°	12.3 76°
$\frac{1}{1+j\omega}$	0 db 0°	−0.6 −21°	−2.2 −39°	−3.8 −50°	−5.4 −57°	−7.0 −63°	−9.4 −70°	−12.3 −76°	−15.7 −81°	−18.1 −83°
$\frac{1}{1-(\omega/2)^2+j\omega/2}$	0 db 0°	0.3 −12°	0.6 −26°	0.9 −46°	1.0 −68°	0 −90°	−4.8 −126°	−12 −148°	−19.5 −160°	−24.5 −166°
Sum = $GH(j\omega)$	20 db 0°	19.9 −22°	19.0 −44°	18.4 −65°	17.8 −86°	16 −108°	10.5 −142°	2.7 −161°	−5.2 −170°	−10.3 −173°

angle of the Bode gain $K_B = 10$ for several frequency values. The db magnitude is 20 db and the phase angle is 0° for all ω. The second row contains the db magnitude and phase angle of the term $(1 + j\omega/2)$ for the same values of ω. These were obtained from Fig. 17-4 by letting $p = 2$ and taking the negatives of the values on the curve for the frequencies in the table. The third row corresponds to the term $1/(1 + j\omega)$ and was also obtained from Fig. 17-4. The fourth row was taken from the $\zeta = 0.5$ curve of Fig. 17-5 by letting $\omega_n = 2$. The sum of the db magnitudes and phase angles of the individual terms for the frequencies in the table is given in the last row. These values are plotted in Fig. 17-6, the db magnitude-phase angle plot of $GH(j\omega)$.

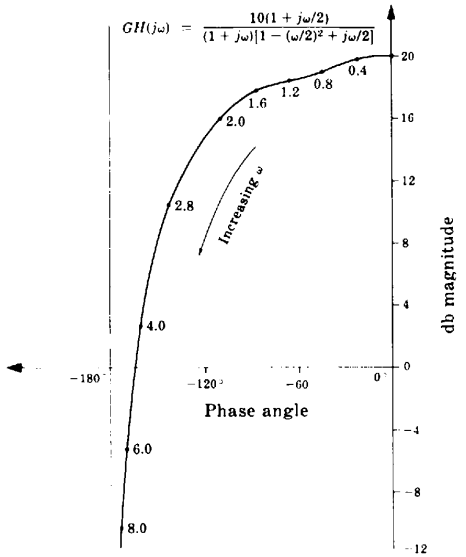


Fig. 17-6

17.4 RELATIVE STABILITY

The gain and phase margins for both continuous-time and discrete-time systems are readily determined from the db magnitude-phase angle plot of $GH(\omega)$.

The *phase crossover* frequency ω_π is the frequency at which the graph of $GH(\omega)$ intersects the $-180°$ line on the db magnitude-phase angle plot. The *gain margin in db* is given by

$$\text{gain margin} = -20\log_{10}|GH(\omega_\pi)| \text{ db} \qquad (17.10)$$

and is read directly from the db magnitude-phase angle plot.

The *gain crossover* frequency ω_1 is the frequency at which the graph of $GH(\omega)$ intersects the 0-db line on the db magnitude-phase angle plot. The phase margin is given by

$$\text{phase margin} = [180 + \arg GH(\omega_1)] \text{ degrees}$$

and can be read directly from the db magnitude-phase angle plot.

In most cases, positive gain and phase margins will ensure stability of the closed-loop system; however, absolute stability should be established by some other means (for example, see Chapters 5 and 11) to guarantee that this is true.

EXAMPLE 17.4. For a stable system, the db magnitude-phase angle plot of $GH(\omega)$ is shown in Fig. 17-7. The gain margin is 15 db and the phase margin is 35°, as indicated.

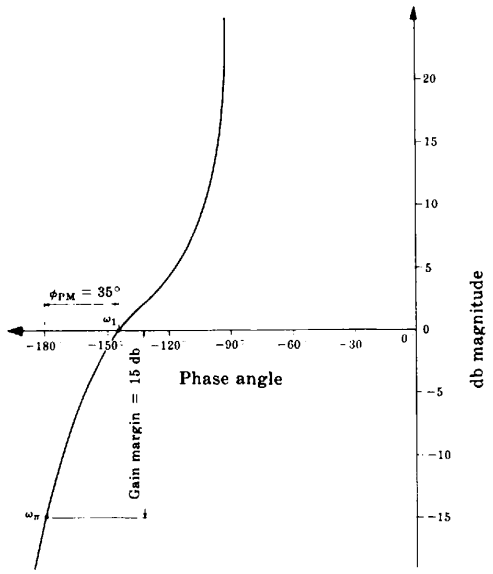


Fig. 17-7

17.5 THE NICHOLS CHART

The remaining discussion is restricted to either continuous-time or discrete-time unity feedback systems. The results are easily generalized to nonunity feedback systems, as illustrated in Example 17.9.

The closed-loop frequency response function of a unity feedback system may be written in polar form as

$$\frac{C}{R}(\omega) = \left|\frac{C}{R}(\omega)\right| \angle \arg\frac{C}{R}(\omega) = \frac{G(\omega)}{1 + G(\omega)} = \frac{|G(\omega)|\angle\phi_G}{1 + |G(\omega)|\angle\phi_G} \qquad (17.11)$$

where $\phi_G \equiv \arg G(\omega)$.

The locus of points on a db magnitude-phase angle plot for which

$$\left|\frac{C}{R}(\omega)\right| = M = \text{constant}$$

is defined by the equation

$$|G(\omega)|^2 + \frac{2M^2}{M^2 - 1}|G(\omega)|\cos\phi_G + \frac{M^2}{M^2 - 1} = 0 \qquad (17.12)$$

For a fixed value of M, this locus can be plotted in three steps: (1) choose numerical values for $|G(\omega)|$; (2) solve the resultant equations for ϕ_G, excluding values of $|G(\omega)|$ for which $|\cos\phi_G| > 1$; and (3) plot the points obtained on a db magnitude-phase angle plot. Note that for fixed values of M and $|G(\omega)|$, ϕ_G is multiple-valued because it appears in the equation as $\cos\phi_G$.

EXAMPLE 17.5. The locus of points for which

$$\left|\frac{C}{R}(\omega)\right| = \sqrt{2}$$

or, equivalently,

$$20\log_{10}\left|\frac{C}{R}(\omega)\right| = 3 \text{ db}$$

is graphed in Fig. 17-8. A similar curve appears at all odd multiples of 180° along the arg $G(\omega)$ axis.

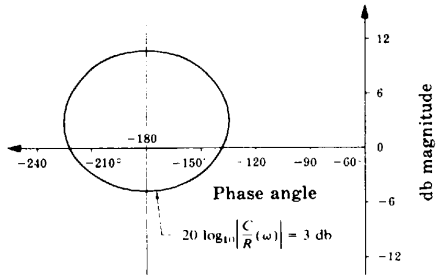


Fig. 17-8

The locus of points on a db magnitude-phase angle plot for which $\arg(C/R)(\omega)$ is constant or, equivalently,

$$\tan\left[\arg\frac{C}{R}(\omega)\right] = N = \text{constant}$$

is defined by the equation

$$|G(\omega)| + \cos\phi_G - \frac{1}{N}\sin\phi_G = 0 \qquad (17.13)$$

For a fixed value of N, this locus of points can be plotted in three steps: (1) choose values for ϕ_G; (2) solve the resultant equations for $G(\omega)$; and (3) plot the points obtained on a db magnitude-phase angle plot.

EXAMPLE 17.6. The locus of points for which $\arg(C/R)(\omega) = -60°$ or, equivalently,

$$\tan\left[\arg\frac{C}{R}(\omega)\right] = -\sqrt{3}$$

is graphed in Fig. 17-9. A similar curve appears at all multiples of 180° along the arg $G(\omega)$ axis.

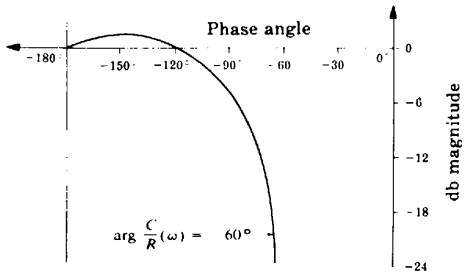


Fig. 17-9

Definition 17.2: A **Nichols chart** is a db magnitude-phase angle plot of the loci of constant db magnitude and phase angle of $(C/R)(\omega)$, graphed as $|G(\omega)|$ versus $\arg G(\omega)$.

EXAMPLE 17.7. A Nichols chart is shown in Fig. 17-10. The range of $\arg G(\omega)$ on this chart is well suited to control system analysis.

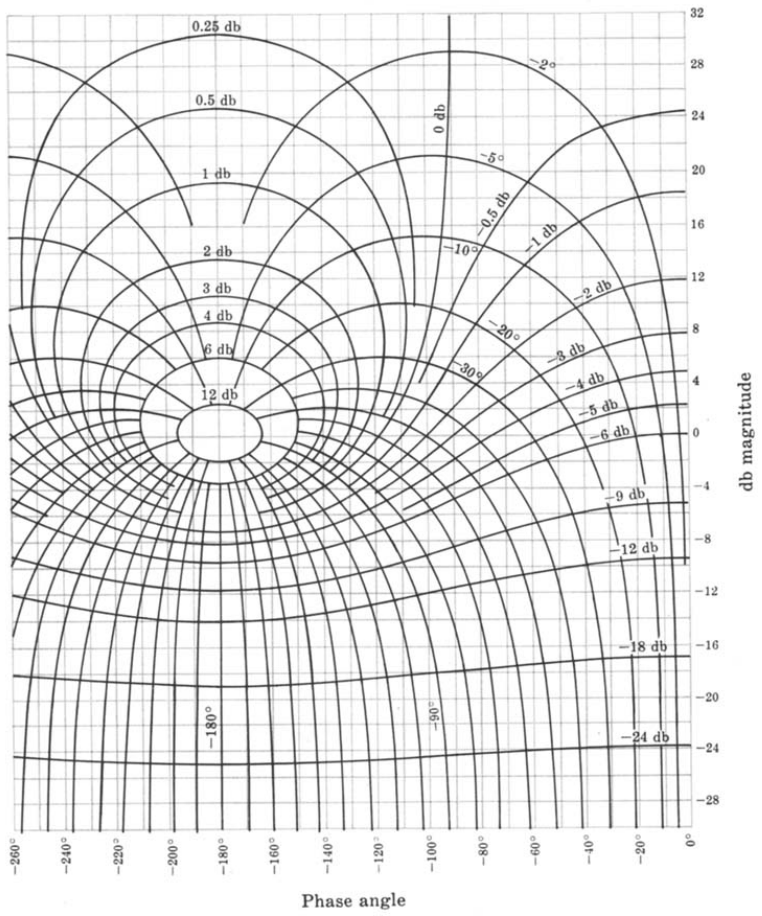


Fig. 17-10

Definition 17.3: A **Nichols chart plot** is a db magnitude-phase angle plot of a frequency response function $P(\omega)$ superimposed on a Nichols chart.

17.6 CLOSED-LOOP FREQUENCY RESPONSE FUNCTIONS

The frequency response function $(C/R)(\omega)$ of a unity feedback system can be determined from the Nichols chart plot of $G(\omega)$. Values of $|(C/R)(\omega)|$ in db and $\arg(C/R)(\omega)$ are determined directly from the plot as the points where the graph of $G(\omega)$ intersects the graphs of loci of constant $|(C/R)(\omega)|$ and $\arg(C/R)(\omega)$.

EXAMPLE 17.8. The Nichols chart plot of $GH(\omega)$ for the continuous-time system of Example 17.3 is shown in Fig. 17-11. Assuming that it is a unity feedback system ($H = 1$), values for $|(C/R)(\omega)|$ and $\arg(C/R)(\omega)$ are obtained from this graph and plotted as a db magnitude-phase angle plot of $(C/R)(\omega)$ in Fig. 17-12.

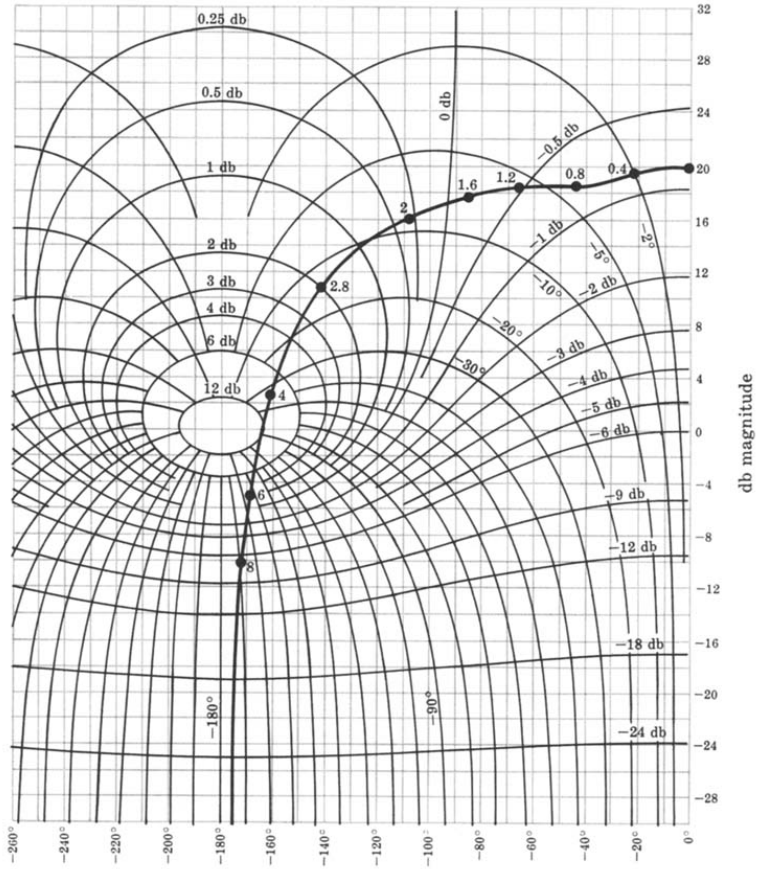


Fig. 17-11

EXAMPLE 17.9. Assume that the system in Example 17.3 is not a unity feedback system and that

$$G(\omega) = \frac{10}{(1+j\omega)\left[1-(\omega/2)^2+j\omega/2\right]} \qquad H(\omega) = 1 + j\frac{\omega}{2}$$

Then

$$\frac{C}{R}(\omega) = \frac{1}{H(\omega)}\left[\frac{GH(\omega)}{1+GH(\omega)}\right] = \frac{1}{H(\omega)}\left[\frac{G'(\omega)}{1+G'(\omega)}\right]$$

where $G' \equiv GH$. The db magnitude-phase angle plot of $G'(\omega)/(1 + G'(\omega))$ was derived in Example 17.8 and is shown in Fig. 17-12. The db magnitude-phase angle plot of $(C/R)(\omega)$ can be obtained by point-by-point addition of the magnitude and phase angle of the pole $1/(1 + j\omega/2)$ to this graph. The magnitude and phase angle of $1/(1 + j\omega/2)$ can be obtained from Fig. 17-4 for $p = 2$. The result is shown in Fig. 17-13.

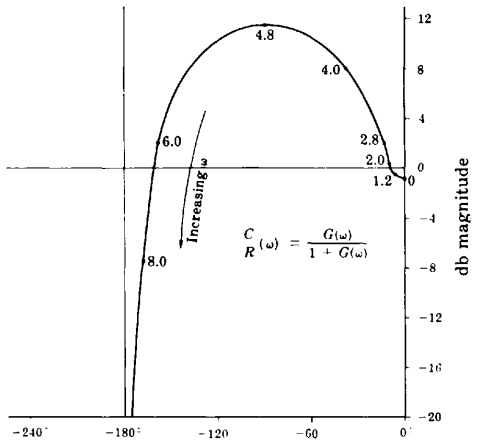


Fig. 17-12

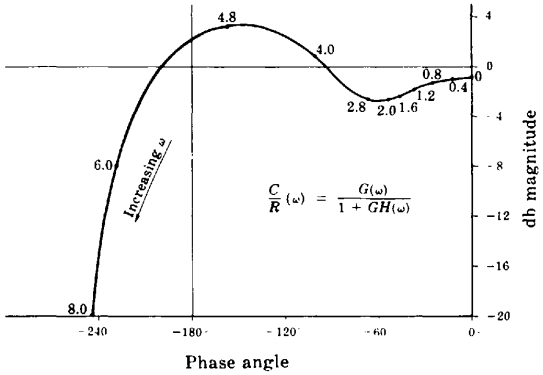


Fig. 17-13

Solved Problems

db MAGNITUDE-PHASE ANGLE PLOTS

17.1. Show that the db magnitude-phase angle plot for a pole of order l at the origin of the s-plane, $1/(j\omega)^l$, is a straight line parallel to the db magnitude axis with an abscissa of $-90l°$ for $\omega \geq 0$.

In polar form, $j\omega = \omega\angle 90°$, $\omega \geq 0$. Therefore

$$\frac{1}{(j\omega)^l} = \frac{1}{\omega^l}\angle -90l° \qquad \omega \geq 0$$

$$20\log_{10}\left|\frac{1}{(j\omega)^l}\right| = 20\log_{10}\frac{1}{\omega^l} = -20\log_{10}\omega^l$$

and $\arg 1/(j\omega)^l = -90l°$. We see that $\arg 1/(j\omega)^l$ is independent of ω; hence the abscissa of the plot is a

constant $-90l°$. In addition, for the region $0 \le \omega \le +\infty$, the db magnitude ranges from $+\infty$ to $-\infty$. Thus the abscissa is fixed and the ordinate takes on all values. The result is a straight line as shown in Fig. 17-3.

17.2. Construct the db magnitude-phase angle plot for the continuous-time open-loop transfer function

$$GH = \frac{2}{s(1+s)(1+s/3)}$$

The db magnitude of $GH(j\omega)$ is

$$20\log_{10}|GH(j\omega)| = 20\log_{10}\frac{2}{|j\omega||1+j\omega||1+j\omega/3|}$$

$$= 20\log_{10}2 - 20\log_{10}\left[\omega\sqrt{1+\omega^2}\sqrt{1+\frac{\omega^2}{9}}\right]$$

$$= 6.02 - 10\log_{10}\left[\omega^2(1+\omega^2)\left(1+\frac{\omega^2}{9}\right)\right]$$

The phase angle of $GH(j\omega)$ is

$$\arg[GH(j\omega)] = -\arg[j\omega] - \arg[1+j\omega] - \arg\left[1+\frac{j\omega}{3}\right]$$

$$= -90° - \tan^{-1}\omega - \tan^{-1}\left(\frac{\omega}{3}\right)$$

The db magnitude-phase angle plot is shown in Fig. 17-14.

17.3. Using the plots in Fig. 17-3 and Fig. 17-4, show how the plot in Fig. 17-14 can be approximated.

We rewrite $GH(j\omega)$ as

$$GH(j\omega) = (2)\left(\frac{1}{j\omega}\right)\left(\frac{1}{1+j\omega}\right)\left(\frac{1}{1+j\omega/3}\right)$$

The db magnitude of $GH(j\omega)$ is

$$20\log_{10}|GH(j\omega)| = 20\log_{10}2 + 20\log_{10}\left|\frac{1}{j\omega}\right| + 20\log_{10}\left|\frac{1}{1+j\omega}\right| + 20\log_{10}\left|\frac{1}{1+j\omega/3}\right|$$

The phase angle is

$$\arg GH(j\omega) = \arg(2) + \arg\left(\frac{1}{j\omega}\right) + \arg\left(\frac{1}{1+j\omega}\right) + \arg\left(\frac{1}{1+j\omega/3}\right)$$

We now construct Table 17.2.

The first row contains the db magnitude and phase angle of the Bode gain $K_B = 2$. The second row contains the db magnitude and phase angle of the term $1/j\omega$ for several values of ω. These are obtained from Fig. 17-3 by letting $l = 1$ and taking values from the curve for the frequencies given. The third row corresponds to the term $1/(1+j\omega)$ and is obtained from Fig. 17-4 for $p = 1$. The fourth row corresponds to the term $1/(1+j\omega/3)$ and is obtained from Fig. 17-4 for $p = 3$. Each pair of entries in the final row is obtained by summing the db magnitudes and phase angles in each column and corresponds to the db magnitude and phase angle of $GH(j\omega)$ for the given value of ω. The values in the last row of this table are then plotted (with the exception of the first) and these points are joined graphically to generate an approximation of Fig. 17-14.

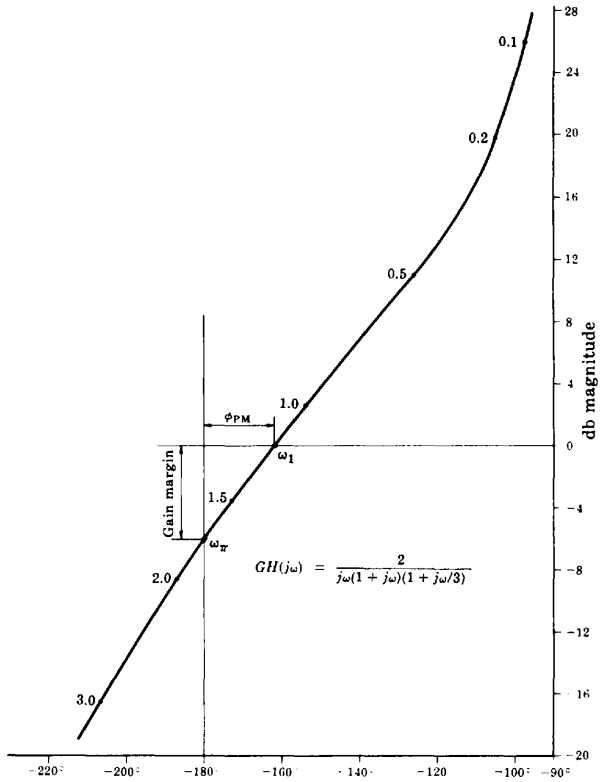


Fig. 17-14

Table 17.2

Term \ Frequency ω	0	0.1	0.2	0.5	1.0	1.5	2.0	3.0
2	6 db 0°	6 0°	6 0°	6 0°	6 0°	6 0°	6 0°	6 0°
$\frac{1}{j\omega}$	∞ $-90°$	20 $-90°$	14 $-90°$	6 $-90°$	0 $-90°$	-3.6 $-90°$	-6 $-90°$	-9.5 $-90°$
$\frac{1}{1+j\omega}$	0 0°	-0.1 $-5.5°$	-0.3 $-11°$	-1.0 $-26°$	-3.0 $-45°$	-5.2 $-57°$	-7.0 $-63°$	-10 $-72°$
$\frac{1}{1+j\omega/3}$	0 0°	0 $-2°$	-0.1 $-4°$	-0.2 $-9°$	-0.5 $-17.5°$	-1.0 $-26°$	-1.6 $-33°$	-3.0 $-45°$
Sum = $GH(j\omega)$	∞ $-90°$	25.9 $-97.5°$	19.6 $-105°$	10.8 $-125°$	2.5 $-152.5°$	-3.8 $-173°$	-8.6 $-186°$	-16.5 $-207°$

17.4. Construct the db magnitude-phase angle plot for the open-loop transfer function

$$GH = \frac{4(s+0.5)}{s^2(s^2+2s+4)}$$

The frequency response function is

$$GH(j\omega) = \frac{4(j\omega+0.5)}{(j\omega)^2((j\omega)^2+2j\omega+4)}$$

A computer-generated db magnitude-phase angle plot of $GH(j\omega)$ is shown in Fig. 17-15.

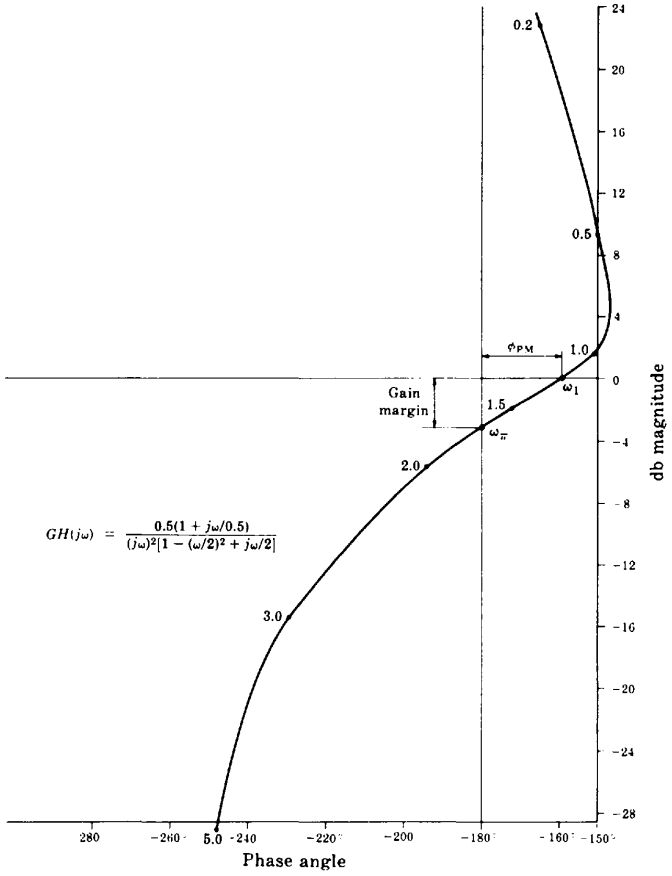


Fig. 17-15

17.5. Construct the db magnitude-phase angle plot for the discrete-time open-loop transfer function

$$GH(z) = \frac{3}{8}\frac{(z+1)(z+\frac{1}{3})}{(z-1)(z+\frac{1}{2})}$$

The open-loop frequency response function is

$$GH(e^{j\omega T}) = \frac{3}{8}\frac{(e^{j\omega T}+1)(e^{j\omega T}+\frac{1}{3})}{(e^{j\omega T}-1)(e^{j\omega T}+\frac{1}{2})}$$

A computer-generated db magnitude-phase angle plot of GH is shown in Fig. 17-16.

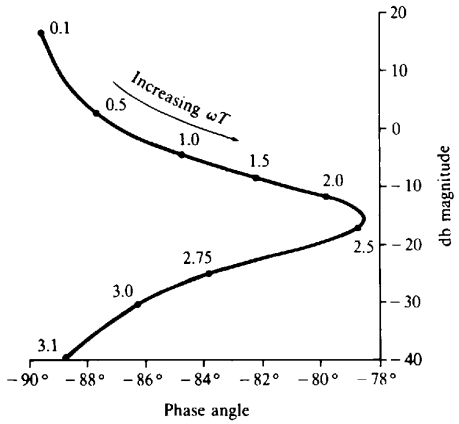


Fig. 17-16

GAIN AND PHASE MARGINS

17.6. Determine the gain and phase margins for the system of Problem 17.2.

The db magnitude-phase angle plot for the open-loop transfer function of this system is given in Fig. 17-14 (Problem 17.2). We see that the curve crosses the 0-db line at a phase angle of $-162°$. Therefore the phase margin is $\phi_{PM} = 180° - 162° = 18°$.

(The gain crossover frequency ω_1 is determined by interpolating along the curve between $\omega = 1.0$ and $\omega = 1.5$ which bound ω_1 below and above, respectively. ω_1 is approximately 1.2 rad/sec.)

The curve crosses the $-180°$ line at a db magnitude of -6 db. Hence gain margin $= -(-6) = 6$ db.

(The phase crossover frequency ω_π is determined by interpolating along the curve between $\omega = 1.5$ and $\omega = 2.0$ which bound ω_π below and above. ω_π is approximately 1.75 rad/sec.)

17.7. Determine the gain and phase margins for the system of Problem 17.4.

The db magnitude-phase angle plot for the open-loop transfer function of this system is given in Fig. 17-15 (Problem 17.4). We see that the curve crosses the 0-db line at a phase angle of $-159°$. Therefore the phase margin is $\phi_{PM} = 180° - 159° = 21°$.

(The gain crossover frequency ω_1 is found by interpolating along the curve between $\omega = 1.0$ and $\omega = 1.5$ which bound ω_1 below and above, respectively. ω_1 is approximately 1.2 rad/sec.)

The curve crosses the $-180°$ line at a db magnitude of -3.1 db. Hence gain margin = 3.1 db.

(The phase crossover frequency ω_π is determined by interpolating between $\omega = 1.5$ and $\omega = 2.0$ which bound ω_π below and above, respectively. ω_π is approximately 1.7 rad/sec.)

17.8. Determine the gain and phase margins for the system defined by the open-loop frequency response function

$$GH(j\omega) = \frac{1+j\omega/0.5}{j\omega[1-(\omega/2)^2+j\omega/2]}$$

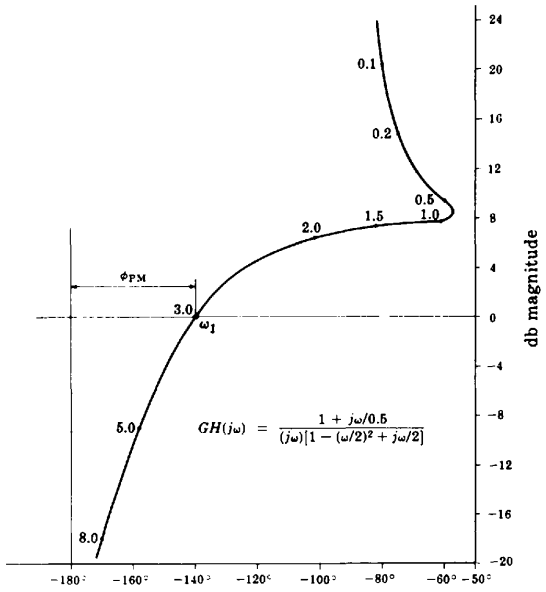


Fig. 17-17

The db magnitude-phase angle plot of $GH(j\omega)$ is given in Fig. 17-17. We see that the curve crosses the 0-db line at a phase angle of $-140°$. Hence the phase margin is $\phi_{PM} = 180° - 140° = 40°$.

The curve does not cross the $-180°$ line for the range of db magnitudes in Fig. 17-17. However, as $\omega \to \infty$,

$$GH(j\omega) \to \frac{j\omega/0.5}{-j\omega(\omega/2)^2} = \frac{8}{\omega^2}\angle -180°$$

The curve approaches the $-180°$ line asymptotically but does not cross it. Therefore the gain margin is indeterminate. This implies that the gain factor can be increased by any amount without producing instability.

17.9. Determine the gain and phase margins for the discrete-time system of Problem 17.5.

The db magnitude-phase angle plot for the open-loop transfer function of this system is given in Fig. 17-16 (Problem 17.5). We see that the curve crosses the 0-db line at a phase angle of $-87°$. Therefore the phase margin $\phi_{PM} = 180° - 87° = 93°$.

The gain crossover angle $\omega_1 T$ can be determined by interpolating along the curve between $\omega T = 0.5$ and $\omega T = 1.0$ which bound $\omega_1 T$ below and above, respectively. $\omega_1 T \simeq 0.6$ rad.

The curve never crosses the $-180°$ line, so the gain margin is indeterminate as is the phase crossover angle.

NICHOLS CHART

17.10. Show that the locus of points on a db magnitude-phase angle plot for which the magnitude of the closed-loop frequency response $(C/R)(\omega)$ of either a continuous-time or discrete-time unity feedback system equals a constant M is defined by Equation (*17.12*).

Using Equation (*17.11*), $|(C/R)(\omega)|$ can be written as

$$\left|\frac{C}{R}(\omega)\right| = \left|\frac{|G(\omega)|\angle\phi_G}{1 + |G(\omega)|\angle\phi_G}\right|$$

Since $|G(\omega)|\angle\phi_G = |G(\omega)|\cos\phi_G + j|G(\omega)|\sin\phi_G$, this can be rewritten as

$$\left|\frac{C}{R}(\omega)\right| = \left|\frac{|G(\omega)|\cos\phi_G + j|G(\omega)|\sin\phi_G}{1 + |G(\omega)|\cos\phi_G + j|G(\omega)|\sin\phi_G}\right|$$

$$= \sqrt{\frac{|G(\omega)|^2\cos^2\phi_G + |G(\omega)|^2\sin^2\phi_G}{[1 + |G(\omega)|\cos\phi_G]^2 + |G(\omega)|^2\sin^2\phi_G}} = \sqrt{\frac{|G(\omega)|^2}{1 + 2|G(\omega)|\cos\phi_G + |G(\omega)|^2}}$$

If we set the last expression equal to M, square both sides, and clear the fraction, we obtain

$$M^2\left[|G(\omega)|^2 + 2|G(\omega)|\cos\phi_G + 1\right] = |G(\omega)|^2$$

which can be written as

$$(M^2 - 1)|G(\omega)|^2 + 2M^2|G(\omega)|\cos\phi_G + M^2 = 0$$

Dividing by $(M^2 - 1)$, we obtain Equation (*17.12*), as required.

17.11. Show that the locus of points on a db magnitude-phase angle plot for which the tangent of the argument of the closed-loop frequency response function $(C/R)(\omega)$ of a unity feedback system equals a constant N is defined by Equation (*17.13*).

Using Equation (*17.11*), $\arg(C/R)(\omega)$ can be written as

$$\arg\left[\frac{C}{R}(\omega)\right] = \arg\left[\frac{|G(\omega)|\angle\phi_G}{1 + |G(\omega)|\angle\phi_G}\right]$$

Since $|G(\omega)|\angle\phi_G = |G(\omega)|\cos\phi_G + j|G(\omega)|\sin\phi_G$,

$$\arg\left[\frac{C}{R}(\omega)\right] = \arg\left[\frac{|G(\omega)|\cos\phi_G + j|G(\omega)|\sin\phi_G}{1 + |G(\omega)|\cos\phi_G + j|G(\omega)|\sin\phi_G}\right]$$

Multiplying numerator and denominator of the term in brackets by the complex conjugate of the denominator yields

$$\arg\left[\frac{C}{R}(\omega)\right] = \arg\left[\frac{(|G(\omega)|\cos\phi_G + j|G(\omega)|\sin\phi_G)(1 + |G(\omega)|\cos\phi_G - j|G(\omega)|\sin\phi_G)}{(1 + |G(\omega)|\cos\phi_G)^2 + |G(\omega)|^2\sin^2\phi_G}\right]$$

Since the denominator of the term in the last brackets is real, $\arg[(C/R)(\omega)]$ is determined by the numerator only. That is,

$$\arg\left[\frac{C}{R}(\omega)\right] = \arg\left[(|G(\omega)|\cos\phi_G + j|G(\omega)|\sin\phi_G)(1 + |G(\omega)|\cos\phi_G - j|G(\omega)|\sin\phi_G)\right]$$

$$= \arg\left[|G(\omega)|\cos\phi_G + |G(\omega)|^2 + j|G(\omega)|\sin\phi_G\right]$$

using $\cos^2\phi_G + \sin^2\phi_G = 1$. Therefore

$$\tan\left[\arg\frac{C}{R}(\omega)\right] = \frac{|G(\omega)|\sin\phi_G}{|G(\omega)|\cos\phi_G + |G(\omega)|^2}$$

Equating this to N, cancelling the common $|G(\omega)|$ term and clearing the fraction, we obtain

$$N\left[\cos\phi_G + |G(\omega)|\right] = \sin\phi_G$$

which can be rewritten in the form of Equation (*17.13*), as required.

17.12. Construct the db magnitude-phase angle plot of the locus defined by Equation (*17.12*) for db magnitude of $(C/R)(\omega)$ equal to 6 db.

$20\log_{10}|(C/R)(\omega)| = 6$ db implies that $|(C/R)(\omega)| = 2$. Therefore we let $M = 2$ in Equation (*17.12*) and obtain

$$|G(\omega)|^2 + \frac{8}{3}|G(\omega)|\cos\phi_G + \frac{4}{3} = 0$$

as the equation defining the locus. Since $|\cos\phi_G| \le 1$, $|G(\omega)|$ may take on only those values for which this constraint is satisfied. To determine bounds of $|G(\omega)|$, we let $\cos\phi_G$ take on its two extreme values of plus and minus unity. For $\cos\phi_G = 1$, the locus equation becomes

$$|G(\omega)|^2 + \frac{8}{3}|G(\omega)| + \frac{4}{3} = 0$$

with solutions $|G(\omega)| = -2$ and $|GH(\omega)| = -\frac{2}{3}$. Since an absolute value cannot be negative, these solutions are discarded. This implies that the locus does not exist on the 0° line (in general, any line which is a multiple of 360°), which corresponds to $\cos\phi_G = 1$.

For $\cos\phi_G = -1$, the locus equation becomes

$$|G(\omega)|^2 - \frac{8}{3}|G(\omega)| + \frac{4}{3} = 0$$

with solutions $|G(\omega)| = 2$ and $|G(\omega)| = \frac{2}{3}$. These are valid solutions for $|G(\omega)|$ and are the extreme values which $|G(\omega)|$ can assume.

Solving the locus equation for $\cos\phi_G$, we obtain

$$\cos\phi_G = \frac{-\left[\frac{4}{3} + |G(\omega)|^2\right]}{\frac{8}{3}|G(\omega)|}$$

The curves obtained from this relationship are periodic with period 360°. The plot is restricted to a single cycle in the vicinity of the −180° line and is obtained by solving for ϕ_G at several values of $|G(\omega)|$ between the bounds 2 and $\frac{2}{3}$. The results are given in Table 17.3

Note that there are two values of ϕ_G whenever $|\cos\phi_G| < 1$. The resulting plot is shown in Fig. 17-18.

Table 17.3

$\lvert G(\omega)\rvert$	$20\log_{10}\lvert G(\omega)\rvert$	$\cos\phi_G$	ϕ_G	
2.0	6 db	−1	−180°	—
1.59	4	−0.910	−204.5°	−155.5°
1.26	2	−0.867	−209.9°	−150.1°
1.0	0	−0.873	−209.2°	−150.8°
0.79	−2	−0.928	−201.9°	−158.1°
0.67	−3.5	−1	−180°	—

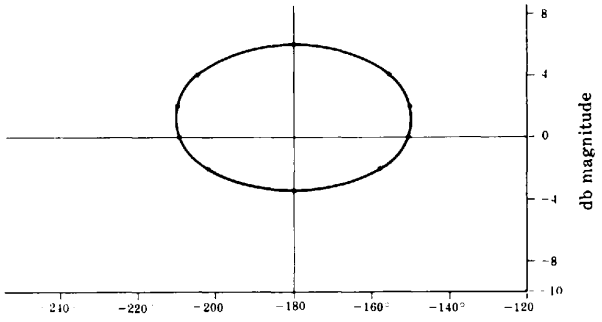


Fig. 17-18

17.13. Construct the db magnitude-phase angle plot of the locus defined by Equation (*17.13*) for $\tan[\arg(C/R)(\omega)] = N = -\infty$,

$\tan[\arg(C/R)(\omega)] = -\infty$ implies that $\arg(C/R)(\omega) = -90 + k360°$, $k = 0, \pm 1, \pm 2, \ldots$, or $\arg(C/R)(\omega) = -270° + k360°$, $k = 0, \pm 1, \pm 2, \ldots$. We will plot only the cycle between −360° and 0°, which corresponds to $k = 0$. Setting $N = -\infty$ in Equation (*17.13*), we obtain the locus equation

$$|G(\omega)| + \cos\phi_G = 0 \qquad \text{or} \qquad \cos\phi_G = -|G(\omega)|$$

Since $|\cos\phi_G| \le 1$, the locus exists only for $0 \le |G(\omega)| \le 1$ or, equivalently,

$$-\infty \le 20\log_{10}|G(\omega)| \le 0$$

To obtain the plot, we use the locus equation to calculate values of db magnitude of $G(\omega)$ corresponding to several values of ϕ_G. The results of these calculations are given in Table 17.4. The desired plot is shown in Fig. 17-19.

Table 17.4

ϕ_G		$\cos\phi_G$	$\lvert G(\omega)\rvert$	$20\log_{10}\lvert G(\omega)\rvert$
−180°	—	−1	1	0 db
−153°	−207°	−0.893	0.893	−1.0
−135°	−222.5°	−0.707	0.707	−3
−120°	−240°	−0.5	0.5	−6
−110.7°	−249.3°	−0.354	0.354	−9
−104.5°	−255.5°	−0.25	0.25	−12
−100.3°	−259.8°	−0.178	0.178	−15

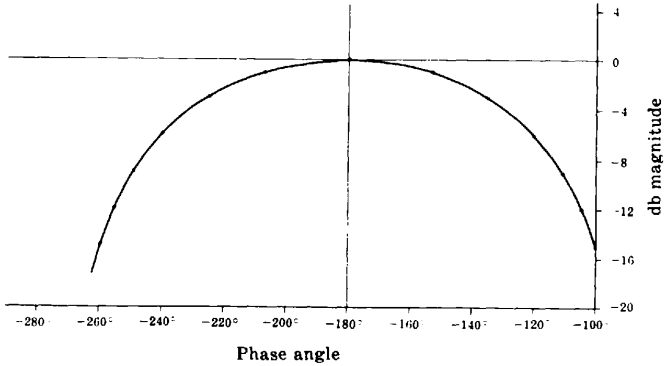


Fig. 17-19

CLOSED-LOOP FREQUENCY RESPONSE FUNCTIONS

17.14. Construct the db magnitude-phase angle plot of the closed-loop frequency response function $(C/R)(j\omega)$ of the unity feedback system whose open-loop transfer function is

$$G = \frac{2}{s(1+s)(1+s/3)}$$

$$\frac{C}{R}(j\omega) = \frac{G(j\omega)}{1+G(j\omega)} = \frac{6}{(j\omega)^3 + 4(j\omega)^2 + 3j\omega + 6} = \frac{6}{(6-4\omega^2) + j(3\omega - \omega^3)}$$

Therefore

$$20\log_{10}\left|\frac{C}{R}(j\omega)\right| = 10\log_{10}\left|\frac{C}{R}(j\omega)\right|^2 = 10\log_{10}\frac{36}{(6-4\omega^2)^2+(3\omega-\omega^3)^2}$$

and

$$\arg\left[\frac{C}{R}(j\omega)\right] = -\tan^{-1}\frac{3\omega-\omega^3}{6-4\omega^2}$$

A computer-generated db magnitude-phase angle plot of $(C/R)(j\omega)$ is shown by the solid line in Fig. 17-20.

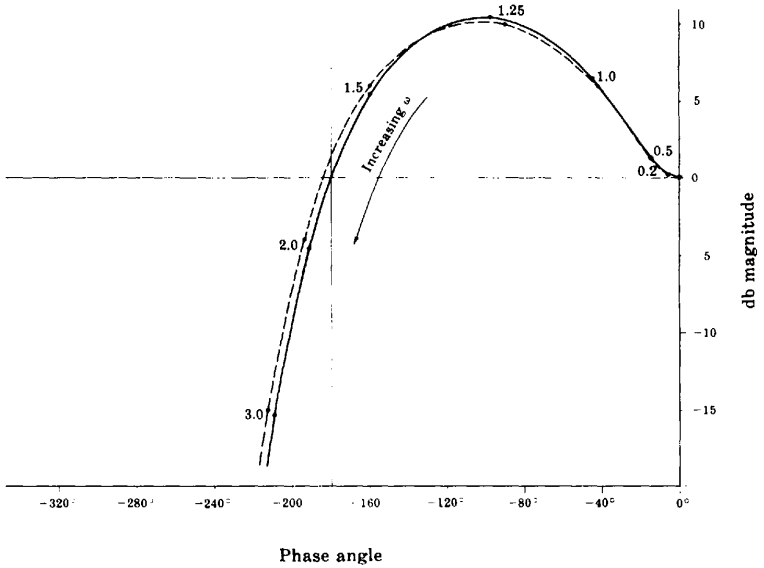


Fig. 17-20

17.15. Solve Problem 17.14 again, using the technique discussed in Section 17.6.

Mathcad

The Nichols chart plot of $G(j\omega)$ is shown in Fig. 17-21. We determine values for the db magnitude of $|(C/R)(j\omega)|$ and $\arg[(C/R)(j\omega)]$ by interpolating values of db magnitude and phase angle on the Nichols chart plot for $\omega = 0, 0.2, 0.5, 1.0, 1.25, 1.5, 2.0, 3.0$. These values are given in Table 17.5.

Table 17.5

ω	$20\log_{10}\left\|\frac{C}{R}(j\omega)\right\|$	$\arg\left[\frac{C}{R}(j\omega)\right]$
0	0 db	0°
0.2	0.2	−6°
0.5	1.2	−15°
1.0	6.0	−42°
1.25	10.0	−90°
1.5	6.0	−155°
2.0	−4.0	−194°
3.0	−15.0	−212°

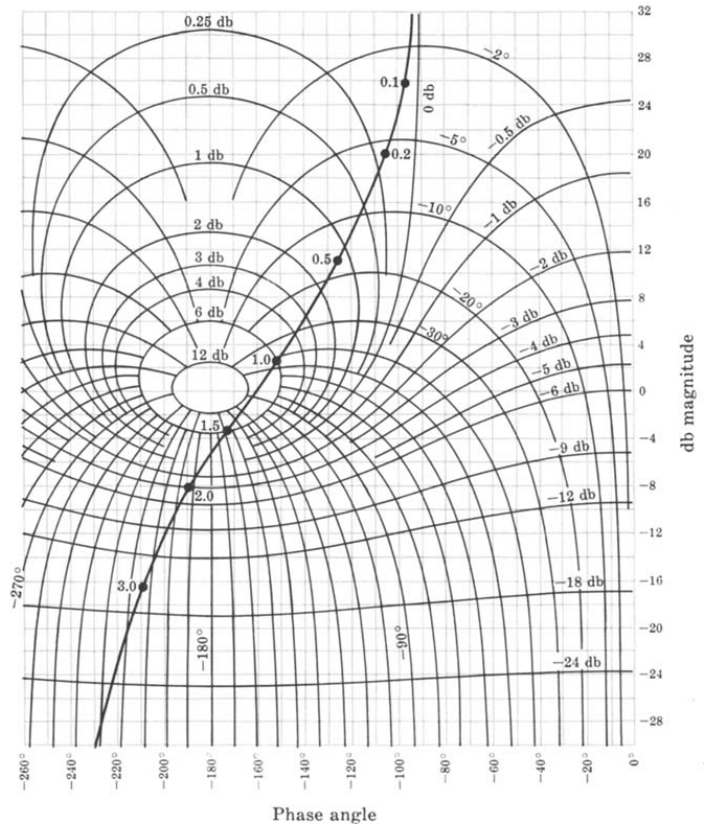


Fig. 17-21

The db magnitude-phase angle plot of $(C/R)(j\omega)$, graphed using the values in the table, is illustrated by the broken line in Fig. 17-20. The differences between the two curves is due to the interpolation necessary to obtain values of db magnitude and phase angle.

Supplementary Problems

17.16. Construct the db magnitude-phase angle plot for the open-loop transfer function

$$GH = \frac{5(s+2)}{s(s+3)(s+5)}$$

17.17. Construct the db magnitude-phase angle plot for the open-loop transfer function

$$GH = \frac{10}{s(1+s/5)(1+s/50)}$$

17.18. Construct the db magnitude-phase angle plot for the open-loop transfer function

$$GH = \frac{1 + s/2}{s(1 + s)(1 + s/4)(1 + s/20)}$$

17.19. Determine gain and phase margins for the system of Problem 17.17.

17.20. Determine the resonance peak M_p and resonant frequency ω_p for the system whose open-loop transfer function is

$$GH = \frac{1}{s(1 + s)(1 + s/4)}$$

17.21. Determine the gain and phase crossover frequencies for the system of Problem 17.17.

17.22. Determine the resonance peak M_p and the resonant frequency ω_p of the system in Problem 17.17.

17.23. Let the system of Problem 17.17 be a unity feedback system and construct the db magnitude-phase angle plot of $(C/R)(j\omega)$.

Answers to Some Supplementary Problems

17.19. Gain margin = 9.5 db, $\phi_{PM} = 25°$

17.20. $M_p = 1.3$ db, $\omega_p = 0.9$ rad/sec

17.21. $\omega_1 = 7$ rad/sec, $\omega_\pi = 14.5$ rad/sec

17.22. $M_p = 8$ db, $\omega_p = 7.2$ rad/sec

Chapter 18

Nichols Chart Design

18.1 DESIGN PHILOSOPHY

Design by analysis in the frequency domain using Nichols chart techniques is performed in the same general manner as the design methods described in previous chapters: appropriate compensation networks are introduced in the forward and/or feedback paths and the behavior of the resulting system is critically analyzed. In this manner, the Nichols chart plot is shaped and reshaped until the performance specifications are met. These specifications are most conveniently expressed in terms of frequency-domain figures of merit such as gain and phase margin for transient performance and the error constants (Chapter 9) for the steady state time-domain response.

The Nichols chart plot is a graph of the open-loop frequency response function $GH(\omega)$, for a continuous-time or discrete-time system, and compensation can be introduced in the forward and/or feedback paths, thus changing $G(\omega)$, $H(\omega)$, or both. We emphasize that no single compensation scheme is universally applicable.

18.2 GAIN FACTOR COMPENSATION

We have seen in several previous chapters (5, 12, 13, 16) that an unstable feedback system can sometimes be stabilized, or a stable system destabilized, by adjustment of the gain factor K of GH. Nichols chart plots are particularly well suited for determining gain factor adjustments. However, when using Nichols techniques for continuous-time systems, it is more convenient to use the Bode gain K_B (Section 15.3), expressed in decibels (db), than the gain factor K. Changes in K_B and K, when given in decibels, are equal.

EXAMPLE 18.1. The db magnitude-phase angle plot for an unstable continuous-time system, represented by $GH(j\omega)$ with the Bode gain $K_B = 5$, is shown in Fig. 18-1. The instability of this system can be verified by a sketch of the Nyquist plot, or application of the Routh criterion. The Nyquist plot in Example 12.1 chapter 12, illustrates the general shape for all Nyquist plots of systems with one pole at the origin and two real poles in the left-half plane. This graph indicates that positive phase and gain margins guarantee stability and negative phase and gain margins guarantee instability for such a system, which implies that a sufficient decrease in the Bode gain stabilizes the system. If the Bode gain is decreased from $20\log_{10}5$ db to $20\log_{10}2$ db, the system is stabilized. The db magnitude-phase angle plot for the compensated system is shown in Fig. 18-2. Further decrease in gain does not alter stability.

Note that the curves for $K_B = 5$ and $K_B = 2$ have identical shapes, the only difference being that the ordinates on the $K_B = 5$ curve exceed those on the $K_B = 2$ curve by $20\log_{10}(5/2)$ db. Therefore changing the gain on a db magnitude-phase angle plot is accomplished by simply shifting the locus of $GH(j\omega)$ up or down by an appropriate number of decibels.

Even though absolute stability can often be altered by gain factor adjustment, this form of compensation is inadequate for most designs because other performance criteria such as those concerned with relative stability cannot usually be met without the inclusion of other types of compensators.

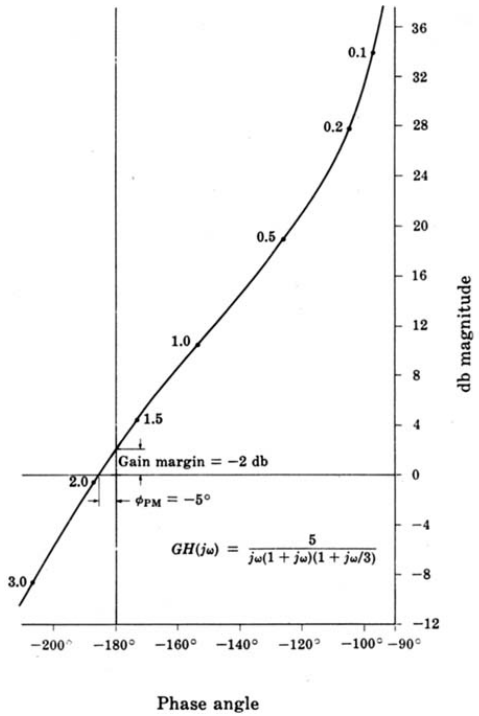


Fig. 18-1

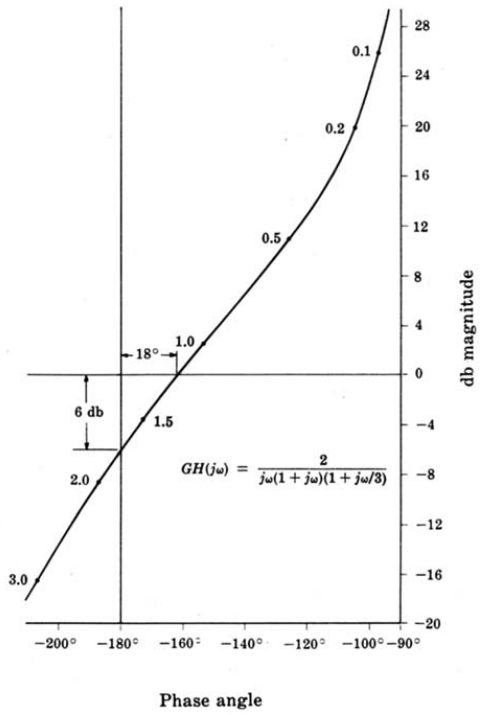


Fig. 18-2

18.3 GAIN FACTOR COMPENSATION USING CONSTANT AMPLITUDE CURVES

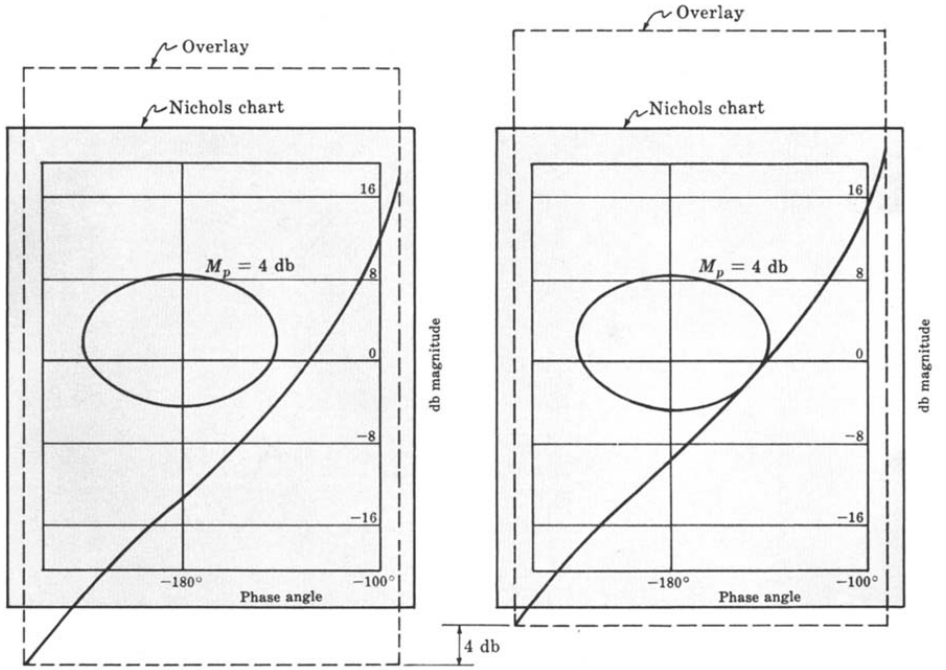
The Nichols chart may be used to determine the gain factor K (for a *unity feedback* system) for a specified resonant peak M_p (in decibels). The following procedure requires drawing the db magnitude-phase angle plot only once.

Step 1: Draw the db magnitude-phase angle plot of $G(\omega)$ for $K = 1$ on tracing paper. The scale of the graph must be the same as that on the Nichols chart.

Step 2: Overlay this plot on the Nichols chart so that the magnitude and phase angle scales of each sheet are aligned.

Step 3: Fix the Nichols chart and slide the plot up or down until it is just tangent to the constant amplitude curve of M_p db. The amount of shift in decibels is the required value of K.

EXAMPLE 18.2. In Fig. 18-3(a), the db magnitude-phase angle plot of the open-loop frequency response function of a particular unity feedback system with $K = 1$ is shown superimposed on a Nichols chart. The desired M_p is 4 db. We see in Fig. 18-3(b) that, if the overlay is shifted upward by 4 db, then the resonant peak M_p of the system is 4 db. Thus the desired K is 4 db.



(a) (b)

Fig. 18-3

18.4 LEAD COMPENSATION FOR CONTINUOUS-TIME SYSTEMS

The Bode form of the transfer function for a lead network is

$$P_{\text{Lead}} = \frac{(a/b)\left(1 + \dfrac{s}{a}\right)}{1 + \dfrac{s}{b}} \qquad (18.1)$$

where $a/b < 1$. The db magnitude-phase angle plots of P_{Lead} for several values of b/a and with the normalized frequency ω/a as the parameter are shown in Fig. 18-4.

For some systems in which lead compensation in the forward loop is applicable, appropriate choice of a and b permits an increase in K_B, providing greater accuracy and less sensitivity, without adversely affecting transient performance. Conversely, for a given K_B, the transient performance can be improved. It is also possible to improve both the steady state and transient responses with lead compensation.

The important properties of a lead network compensator are its phase lead contribution in the low-to-medium-frequency range (the vicinity of the resonant frequency ω_p) and its negligible attenuation at high frequencies. If a very large phase lead is required, several lead networks may be cascaded.

Lead compensation generally increases the bandwidth of a system.

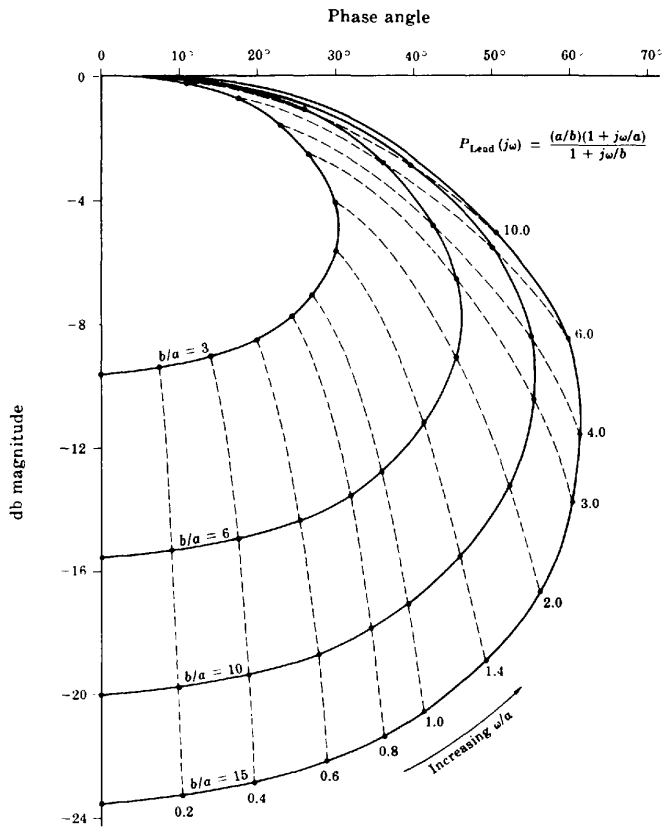


Fig. 18-4

EXAMPLE 18.3. The uncompensated continuous-time unity feedback system whose open-loop transfer function is

$$GH = \frac{2}{s(1+s)(1+s/3)}$$

is to be designed to meet the following performance specifications:

1. When the input is a unit ramp function, the steady state position error must be less than 0.25.
2. $\phi_{PM} \cong 40°$.
3. Resonance peak $\cong$ 4 db.

Note that the Bode gain is equal to the velocity error constant K_v. Therefore the steady state error for the uncompensated system is $e(\infty) = 1/K_v = \frac{1}{2}$ [Equation (9.13)]. From the db magnitude-phase angle plot of GH in Fig. 18-5, we see that $\phi_{PM} = 18°$ and $M_p = 11$ db.

The steady state error is too large by a factor of 2; therefore the Bode gain must be increased by a factor of 2 (6 db). If we increase the Bode gain by 6 db, we obtain the plot labeled GH_1 in Fig. 18-5. The phase margin of GH_1 is about zero and the resonant peak is near infinity. Therefore the system is on the verge of instability.

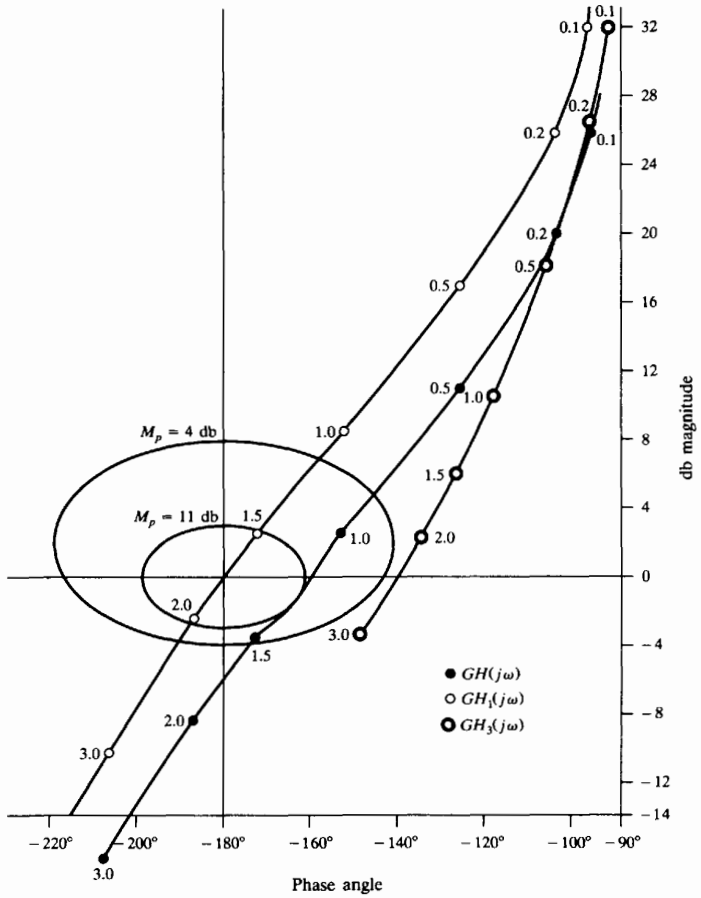


Fig. 18-5

Phase lead compensation can be used to improve the relative stability of the system. The compensated open-loop transfer function is

$$GH_2 = \frac{K_B(a/b)(1+s/a)}{s(1+s)(1+s/3)(1+s/b)} = \frac{4(1+s/a)}{s(1+s)(1+s/3)(1+s/b)}$$

where $K_B = 4(b/a)$ to satisfy the steady state error.

One way of satisfying the requirements on ϕ_{PM} and M_p is to add 40° to 50° of phase lead to the GH_1 curve in the region $1 \le \omega \le 2.5$ without substantially changing the db magnitude. We have already chosen $K_B = 4(b/a)$ to compensate for a/b in the lead network. Therefore we need concern ourselves only with the effect that the factor $(1 + s/a)/(1 + s/b)$ has on the GH_1 curve. Referring to Fig. 18-4, we see that in order to provide the necessary phase lead we will require $b/a \ge 10$. We note that the curves of Fig. 18-4 include the effect of a/b of the lead network. Since we have already compensated for this, we must add 20 $\log_{10}(b/a)$ to the db magnitudes on the curve. In order to keep the db magnitude contribution of the lead network small in the region $1 \le \omega \le 2.5$, we let $b/a = 15$ and choose a so that only the lower portion of the curve ($\omega/a \le 3.0$) contributes in the region of interest $1 \le \omega \le 2.5$. In particular, we let $a = 1.333$. Then the compensated open-loop transfer function is

$$GH_3 = \frac{4(1+s/1.333)}{s(1+s)(1+s/3)(1+s/20)}$$

The db magnitude-phase angle plot of GH_3 is shown in Fig. 18-5. We see that $\phi_{PM} = 40.5°$ and $M_p = 4$ db. Thus the specifications are all met. We note, however, that the resonant frequency ω_p of the compensated system is about 2.25 rad/sec. For the uncompensated system defined by GH it is about 1.2 rad/sec. Thus the bandwidth has been increased.

A block diagram of the fully compensated system is shown in Fig. 18-6.

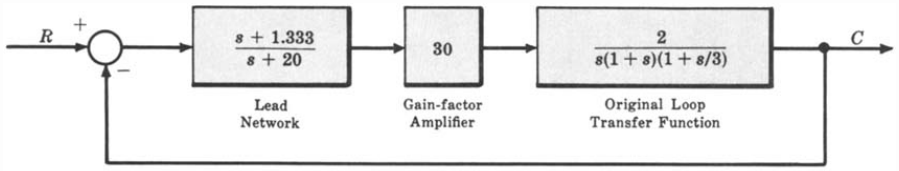


Fig. 18-6

18.5 LAG COMPENSATION FOR CONTINUOUS-TIME SYSTEMS

The Bode form transfer function for a lag network is

$$P_{\text{Lag}} = \frac{1 + s/b}{1 + s/a} \tag{18.2}$$

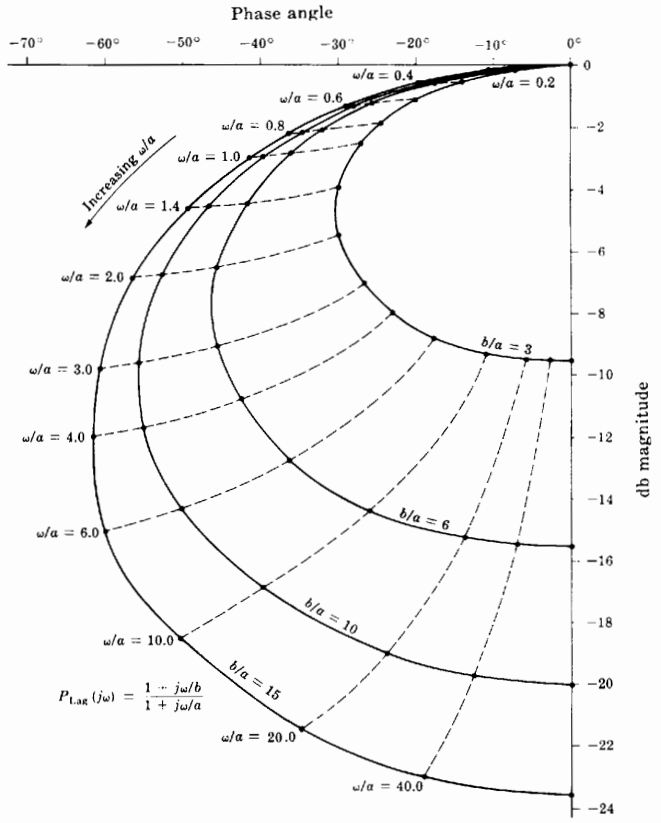


Fig. 18-7

where $a < b$. The db magnitude-phase angle plots of P_{Lag} for several values of b/a and with the normalized frequency ω/a as the parameter are shown in Fig. 18-7.

The lag network provides compensation by attenuating the high-frequency portion of the db magnitude-phase angle plot. Higher attenuation is provided by cascading several lag networks.

Several general effects of lag compensation are:

1. The bandwidth of the system is usually decreased.
2. The dominant time constant τ of the system is usually increased, producing a more sluggish system.
3. For a given relative stability, the value of the error constant is increased.
4. For a given error constant, relative stability is improved.

The procedure for using lag compensation is essentially the same as that for lead compensation.

EXAMPLE 18.4. Let us redesign the system of Example 18.3 using gain factor plus lag compensation. The steady state specification is again satisfied by GH_1. The db magnitude-phase angle plot of GH_1 is repeated in Fig. 18-8. Since $P_{\text{Lag}}(j0) = 1$, introduction of the lag network after the steady state specification has been met by gain factor compensation does not require an additional increase in gain factor.

Incorporating the lag network, we get the open-loop transfer function

$$GH_4 = \frac{4(1 + s/b)}{s(1 + s)(1 + s/3)(1 + s/a)}$$

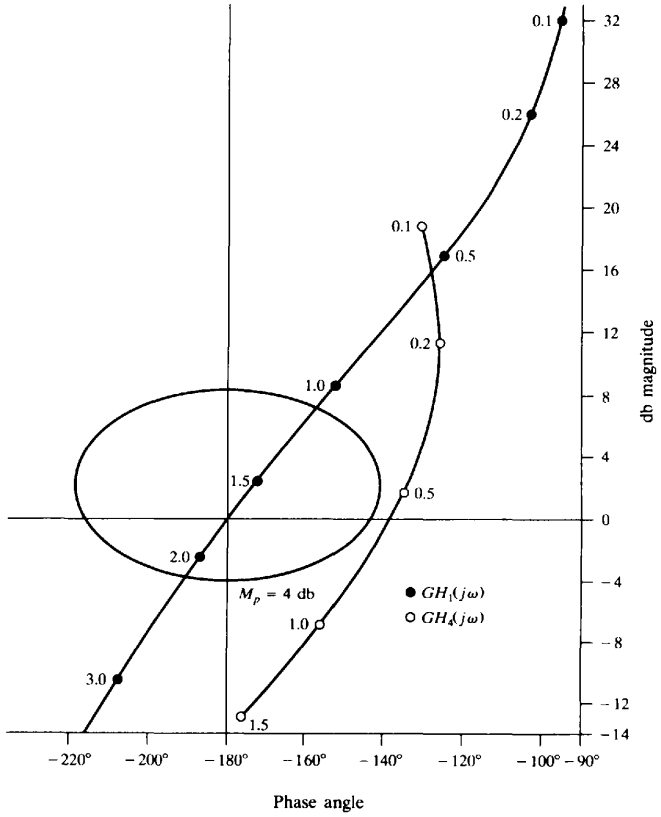


Fig. 18-8

One way of satisfying the requirements on ϕ_{PM} and M_p is to choose a and b such that the GH_1 curve is attenuated by about 12 db in the region $0.7 \leq \omega \leq 2.0$ without substantial change in the phase angle. Since the lag network introduces some phase lag, it is necessary to attenuate the curve more than 12 db. Referring to Fig. 18-7, we see that if we choose $b/a = 6$, a maximum of 15.5-db attenuation is possible. If we choose $a = 0.015$, then at a frequency $\omega = 0.5$ ($\omega/a = 33.33$) 15.4 db of attenuation is obtained from the lag network, with a phase lag of $-9°$. GH_4 can now be written as

$$GH_4 = \frac{4(1 + s/0.09)}{s(1+s)(1+s/3)(1+s/0.015)}$$

where $b = 6a = 0.09$. The db magnitude-phase angle plot of GH_4 is given in Fig. 18-8. We see that $\phi_{PM} = 41°$ and $M_p \cong 4$, which satisfy the specifications. We note that the resonant frequency ω_p of the compensated system is about 0.5 rad/sec. For the uncompensated system defined by GH, ω_p is about 1.2 rad/sec. A block diagram of the fully compensated system is shown in Fig. 18-9.

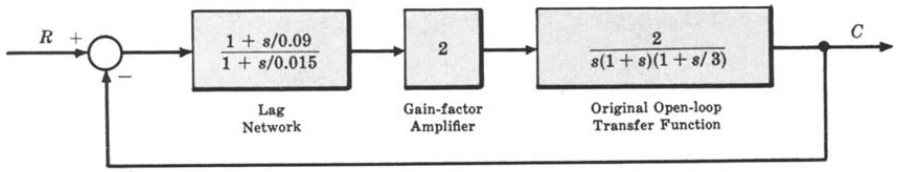


Fig. 18-9

18.6 LAG-LEAD COMPENSATION

The Bode form transfer function for a lag-lead network is

$$P_{LL} = \frac{(1 + s/a_1)(1 + s/b_2)}{(1 + s/b_1)(1 + s/a_2)} \qquad (18.3)$$

where $b_1/a_1 = b_2/a_2 > 1$. The db magnitude-phase angle plots of P_{LL} for a few values of b_1/a_1 ($= b_2/a_2$), when $a_1/a_2 = 6, 10, 100$, and with the normalized frequency ω/a_2 are shown in Fig. 18-10(a), (b), and (c)

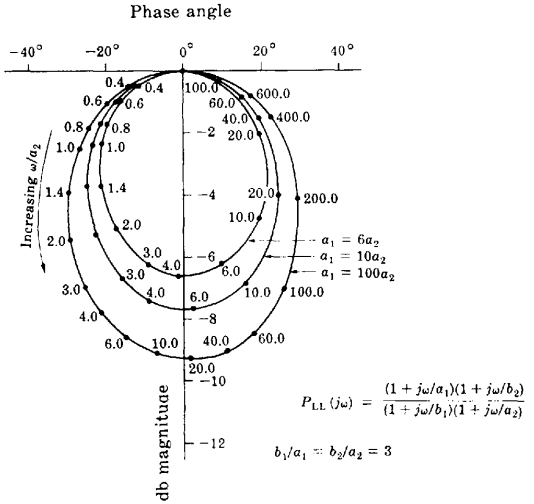


Fig. 18-10(a)

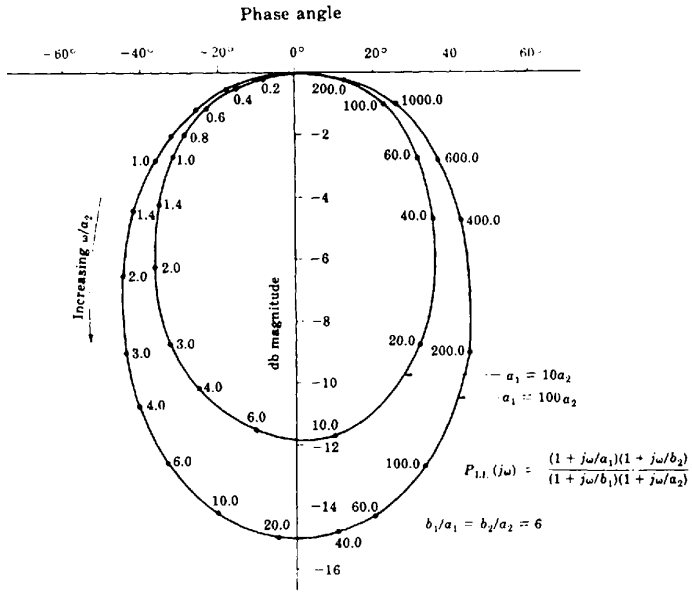


Fig. 18-10(b)

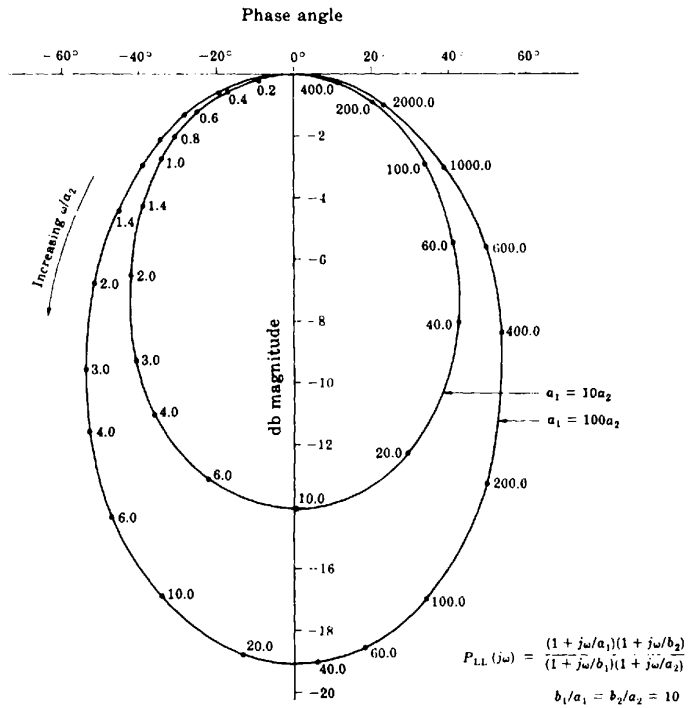


Fig. 18-10(c)

Additional plots of P_{LL} for other values of b_1/a_1 and a_1/a_2 can be obtained by combining plots of lag networks (Fig. 18-7) and lead networks (Fig. 18-4).

Lag-lead compensation has all of the advantages of both lag and lead compensation and a minimum of their usually undesirable characteristics. For example, system specifications can be satisfied without excessive bandwidth or sluggish time response caused by phase lead or lag, respectively.

EXAMPLE 18.5. Let us redesign the system of Example 18.3 using gain factor plus lag-lead compensation. We add the additional specification that the resonant frequency ω_p of the compensated system must be approximately the same as that of the uncompensated system. The steady state specification is again satisfied by

$$GH_1 = \frac{4}{s(1+s)(1+s/3)}$$

as shown in Example 18.3. Since $P_{LL}(j0) = 1$, introduction of the lag-lead network does not require an additional increase in gain factor.

Inserting the lag-lead network, we get the open-loop transfer function

$$GH_5 = \frac{4(1+s/a_1)(1+s/b_2)}{s(1+s)(1+s/3)(1+s/b_1)(1+s/a_2)}$$

From Fig. 18-5, we see that for the uncompensated system GH, $\omega_p = 1.2$ rad/sec. From the db magnitude-phase angle plot of GH_1 (Fig. 18-11) we see that, if $GH_1(j1.2)$ is attenuated by 6.5 db and has its phase increased by 20°, the resonant frequency $\omega_p = 1.2$ is shifted to $M_p = 4$ db. Referring to Fig. 18-10(*a*), we see that the desired attenuation and phase lead are obtained with $b_1/a_1 = b_2/a_2 = 3$, $a_1/a_2 = 10$, and $\omega/a_2 = 12$. The constants a_1,

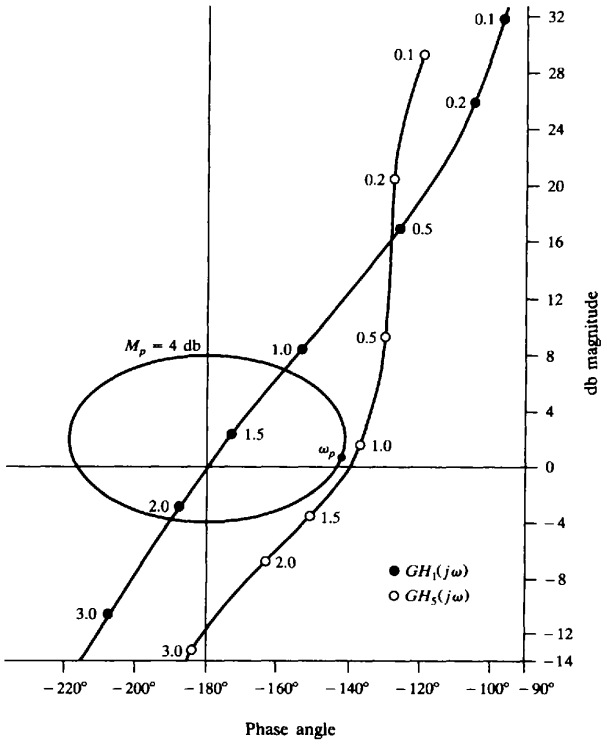


Fig. 18-11

a_2, b_1, and b_2 are determined by noting that

$$a_2 = \frac{\omega_p}{12} = \frac{1.2}{12} = 0.1 \qquad a_1 = 10a_2 = 1 \qquad b_2 = 3a_2 = 0.3 \quad \text{and} \quad b_1 = 3a_1 = 3$$

GH_5 then becomes

$$GH_5 = \frac{4(1+s)(1+s/0.3)}{s(1+s)(1+s/3)(1+s/3)(1+s/0.1)} = \frac{4(1+s/0.3)}{s(1+s/3)^2(1+s/0.1)}$$

The complete db magnitude-phase angle plot of GH_5 is shown in Fig. 18-11. We see that $\phi_{PM} = 40.5°$, $M_p = 4$ db, and the resonant frequency $\omega_p \cong 1.15$. Thus all specifications have been satisfied.

8.7 NICHOLS CHART DESIGN OF DISCRETE-TIME SYSTEMS

As with Bode methods (Section 16.6), design of discrete-time systems using Nichols charts is not as straightforward as the design of continuous-time systems using either of these approaches. But, again, the w-transform can facilitate the process as it did for Bode design of discrete-time systems. The method is the same as that developed in Section 16.6.

EXAMPLE 18.6. The uncompensated discrete-time unity feedback system with plant transfer function

$$G_2(z) = \frac{9}{4}\frac{(z+1)^3}{z\left(z+\frac{1}{2}\right)^2}$$

is to be designed to yield an overall phase margin of 40° and the same gain crossover frequency ω_1 as the uncompensated system. Since both of these specifications are in the frequency domain, we transform the problem directly into the w-domain by substituting $z = (1+w)/(1-w)$, thus forming

$$G_2'(w) = \frac{72}{(w+1)(w+3)^2}$$

The db magnitude-phase angle plot for this system is shown in Fig. 18-12. The gain crossover frequency obtained from this plot is $\omega_{w1} = 3.4$ rad/sec and the phase margin is 10°. A lead compensator with somewhat arbitrary a and b can be chosen as long as the phase lead at $\omega_{w1} = 3.4$ rad/sec is sufficient to raise the phase margin from 10° to 40°. The minimum b/a ratio that yields 30° of phase lead is about 3.3 from Fig. 18-4. We choose a and b so that the maximum phase lead occurs at $\omega_{w1} = 3.4$ rad/sec. From Section 16.3, this occurs when $\omega_{w1} = 3.4 = \sqrt{ab}$. Since $b = 3.3a$, we find $b = 6.27$ and $a = 1.90$. This compensator produces about $20\log_{10}\sqrt{6.27/1.90} = 5$ db of attenuation at $\omega_{w1} = 3.4$ rad/sec. Thus an amplifier with gain of 5.2 db, or gain factor 1.82, is required in addition

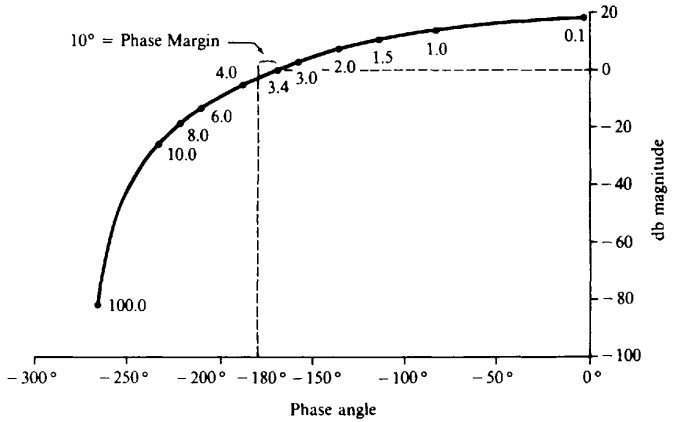


Fig. 18-12

to the lead compensator to maintain ω_{w1} at 3.4 rad/sec. The w-domain transfer function for the compensator is therefore given by

$$G_1(w) = \frac{1.82(w + 1.90)}{w + 6.27}$$

This is transformed back to the z-domain by letting $w = (z-1)/(z+1)$, thus forming

$$G_1(z) = \frac{0.7229(z + 0.3007)}{z + 0.7222}$$

The compensated control system is shown in Fig. 18-13.

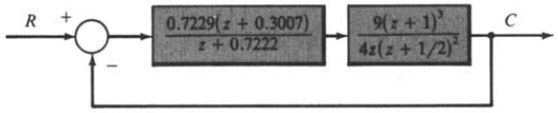


Fig. 18-13

Solved Problems

GAIN FACTOR COMPENSATION

18.1. The db magnitude-phase angle plot of the open-loop continuous-time frequency response function

$$GH(j\omega) = \frac{K_B\left[1 - (\omega/2)^2 + j\omega/2\right]}{j\omega(1 + j\omega/0.5)^2(1 + j\omega/4)}$$

is shown in Fig. 18-14 for $K_B = 1$. The closed-loop system defined by $GH(j\omega)$ is stable for $K_B = 1$. Determine a value of K_B for which the phase margin is 45°.

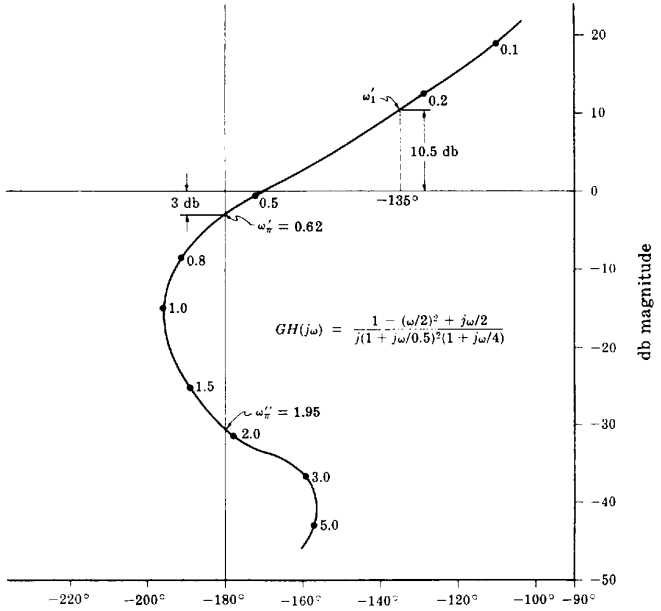


Fig. 18-14

$\phi_{PM} = 180° + \arg GH(j\omega_1)$, where ω_1 is the gain crossover frequency. For $\phi_{PM} = 45°$, ω_1 must be chosen so that $\arg GH(j\omega_1) = -135°$. If we draw a vertical line with abscissa of $-135°$, it intersects the $GH(j\omega)$ curve at a point $\omega_1' \cong 0.25$ rad/sec, where $\arg GH(j\omega_1') = -135°$. The ordinate of this point of intersection is 10.5 db. If we decrease K_B by 10.5 db, the gain crossover frequency becomes ω_1', and $\phi_{PM} = 45°$. A decrease of 10.5 db implies that $20\log_{10} K_B = -10.5$, or $K_B = 10^{-10.5/20} = 0.3$. Further decrease in K_B increases ϕ_{PM} beyond 45°.

18.2. For the system in Problem 18.1, determine the value of K_B for which the system is stable and the gain margin is 10 db.

Gain margin $= -20\log_{10}|GH(j\omega_\pi)|$ db, where ω_π is the phase crossover frequency. Referring to Fig. 18-14, we see that there are two phase crossover frequencies: $\omega_\pi' \cong 0.62$ rad/sec and $\omega_\pi'' \cong 1.95$ rad/sec. For $\omega_\pi' = 0.62$, we have $20\log_{10}|GH(j\omega_\pi')| = -3$ db. Therefore the gain margin is 3 db. It can be increased to 10 db by shifting the $GH(j\omega)$ curve downward by 7 db. The phase crossover frequency ω_π' is the same in the new position, but $20\log_{10}|GH(j\omega_\pi')| = -10$ db. A gain decrease of 7 db implies that $K_B = 10^{-7/20} = 0.447$. Since the system is stable for $K_B = 1$, it remains stable when the $GH(j\omega)$ curve is shifted downward. Absolute stability is not affected unless the $GH(j\omega)$ curve is shifted upward and across the point defined by 0 db and $-180°$, as would be necessary if $-20\log_{10} GH(j\omega_\pi'') = 10$ db.

18.3. For the system of Problem 18.1, determine a value for K_B such that: gain margin ≥ 10 db, $\phi_{PM} \geq 45°$.

In Problem 18.1, it was shown that $\phi_{PM} \geq 45°$ if $K_B \leq 0.3$; in Problem 18.2, gain margin ≥ 10 db if $K_B \leq 0.447$. Therefore both requirements can be satisfied by setting $K_B \leq 0.3$. Note that if we had specified gain margin = 10 db and $\phi_{PM} = 45°$, then the specifications could not be met by gain factor compensation alone.

18.4. Assume that the system of Problem 18.1 is a unity feedback system and determine a value for K_B such that the resonant peak M_p is 5 db.

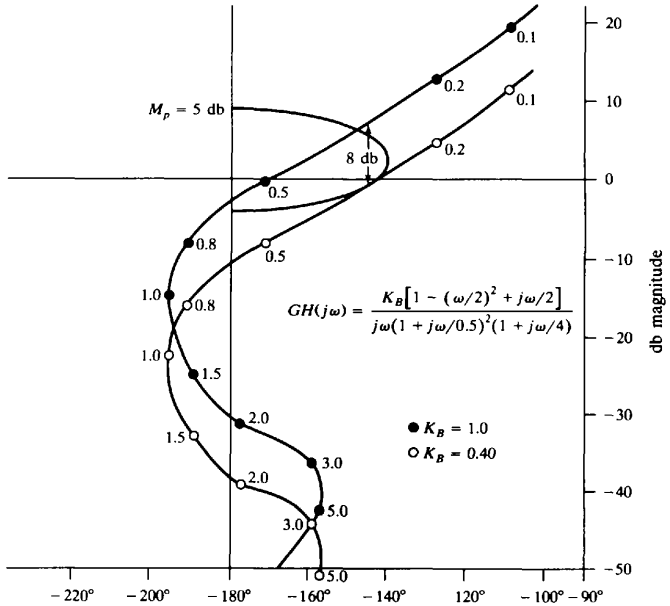


Fig. 18-15

The db magnitude-phase angle plot of $GH(j\omega)$ for $K_B = 1$ is shown in Fig. 18-15 along with the locus of points for which $|(C/R)(j\omega)| = 2$ db ($M_p = 2$ db). We see that if K_B is decreased by 8 db, the resulting $GH(j\omega)$ curve is just tangent to the $M_p = 2$ db curve. A decrease of 8 db implies that $K_B = 10^{-8/20} = 0.40$.

18.5. The db magnitude-phase angle plot of the open-loop frequency response function

$$GH(j\omega) = \frac{K_B(1 + j\omega/0.5)}{(j\omega)^2\left[1 - (\omega/2)^2 + j\omega/2\right]}$$

is given in Fig. 18-16 for $K_B = 0.5$. The closed-loop system defined by $GH(j\omega)$ is stable for $K_B = 0.5$. Determine the value of K_B which maximizes the phase margin.

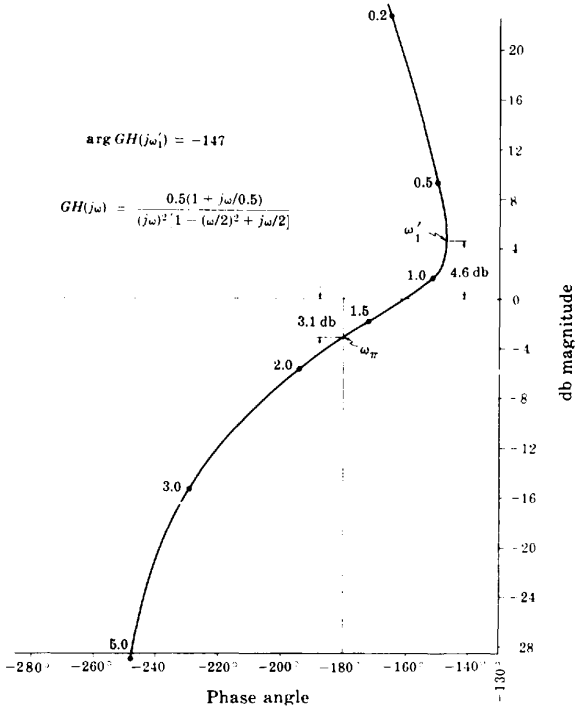


Fig. 18-16

$\phi_{PM} = 180° + \arg GH(j\omega_1)$, where ω_1 is the gain crossover frequency. Referring to Fig. 18-16, we see that $\arg GH(j\omega)$ is always negative. Therefore if we maximize $\arg GH(j\omega_1)$, ϕ_{PM} will be maximized. Fig. 18-16 indicates that $\arg GH(j\omega)$ is maximum when $\omega = \omega_1' \cong 0.8$ rad/sec and $\arg GH(j\omega_1') = -147°$. The ordinate of the point $GH(j\omega_1')$ is 4.6 db. Therefore if K_B is decreased by 4.6 db, the phase crossover frequency becomes ω_1'; and ϕ_{PM} takes on its maximum value: $\phi_{PM} = 180° + \arg GH(j\omega_1') = 33°$. A decrease of 4.6 db in K_B implies that $20\log_{10}(K_B/0.5) = -4.6$ db or $K_B/0.5 = 10^{-4.6/20}$. Then $K_B = 0.295$.

18.6. For the system in Problem 18.5, determine a value of K_B for which the system is stable and the gain margin is 8 db.

Gain margin $= -20\log_{10}|GH(j\omega_\pi)|$ db. Referring to Fig. 18-16, we see that the gain margin is 3.1 db. This can be increased to 8 db by shifting the curve down by 4.9 db; ω_π remains the same, as it is independent of K_B. A decrease of 4.9 db in K_B implies that $20\log_{10}(K_B/0.5) = -4.9$ or $K_B = 0.254$.

PHASE COMPENSATION

18.7. The db magnitude-phase angle plot of the open-loop transfer function $G(j\omega)$ for a particular unity feedback system has been determined experimentally as shown in Fig. 18-17. In addition, the steady state error $e(\infty)$ for a unit ramp function input was measured and found to be $e(\infty) = 0.2$. The open-loop transfer function is known to have a pole at the origin. Determine a combination of phase lead plus gain compensation such that: $M_p \cong 3.5$ db, $\phi_{PM} \cong 40°$, and the steady state error for a unit ramp input is $e(\infty) = 0.1$.

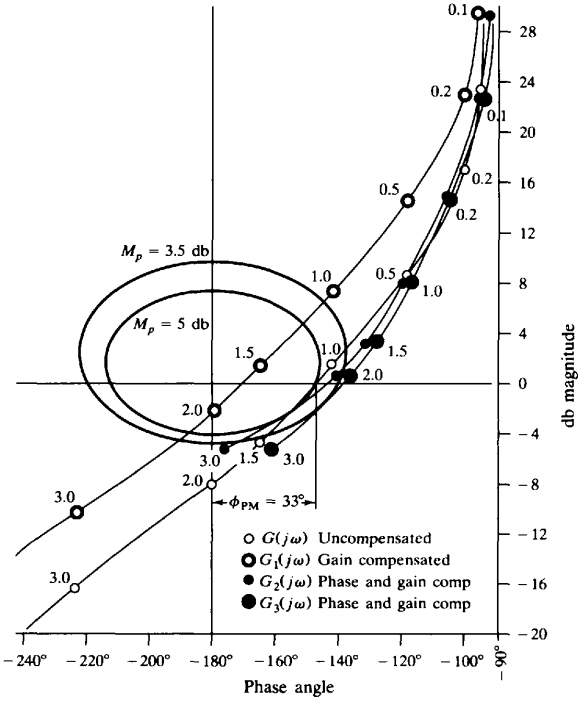


Fig. 18-17

Since $e(\infty) = 1/K_v = 1/K_B$, the steady state requirement can be satisfied by doubling K_B. The compensation has the form

$$K'P_{Lead}(j\omega) = \frac{K'(a/b)(1 + s/a)}{1 + s/b}$$

Hence K_B is doubled by letting $K'(a/b) = 2$, or $K' = 2(b/a)$.

The db magnitude-phase angle plot for the gain compensated open-loop frequency response function

$$G_1(j\omega) = 2G(j\omega)$$

is shown in Fig. 18-17. $G_1(j\omega)$ satisfies the steady state specification. To satisfy the specifications on M_p and ϕ_{PM}, the $G_1(j\omega)$ curve must be shifted to the right by about 30° to 40° in the region $1.2 \le \omega \le 2.5$

without substantially changing the db magnitude. This is done by proper choice of a and b. Referring to Fig. 18-4, we see that, for $b/a = 10$, 30° phase lead is obtained for $\omega/a \geq 0.65$. Since the lead ratio a/b of the lead network is taken into account by designing for the gain factor $K' = 2(b/a) = 20$, we must add $20\log(b/a) = 20\log_{10}10 = 20$ db to all db magnitudes taken from Fig. 18-4.

To obtain 30° or more phase lead in the frequency range of interest, we let $a = 2$. For this choice we have $\omega = (2)(0.65) = 1.3$ and obtain 30° phase lead. Since $b/a = 10$, then $b = 20$. The compensated open-loop frequency response function is

$$G_2(j\omega) = \frac{2(1+j\omega/2)}{1+j\omega/20}G(j\omega)$$

The db magnitude-phase angle plot of $G_2(j\omega)$ is shown in Fig. 18-17. We see that $M_p \cong 4.0$ db and $\phi_{PM} = 36°$; therefore the specifications are not satisfied by this compensation. We need to shift $G_2(j\omega)$ 5° to 10° further to the right; hence additional phase lead is needed. Referring once more to Fig. 18-4, we see that letting $b/a = 15$ increases the phase lead. Again, we let $a = 2$; then $b = 30$. The db magnitude-phase angle plot of

$$G_3(j\omega) = \frac{2(1+j\omega/2)}{1+j\omega/30}G(j\omega)$$

is shown in Fig. 18-17. We see that $\phi_{PM} = 41°$ and $M_p \cong 3.5$ db and hence the specifications are met by the compensation

$$30P_{Lead} = \frac{2(1+s/2)}{1+s/30}$$

18.8. Solve Problem 18.7 using *lag* plus gain compensation.

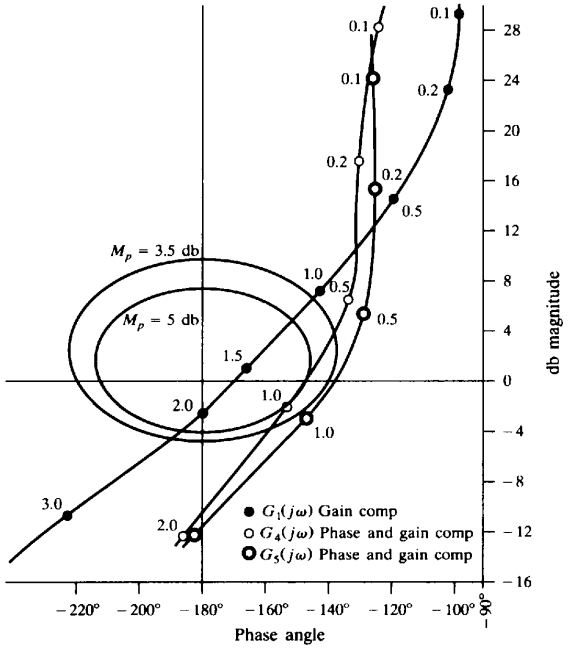


Fig. 18-18

In Problem 18.7 we found that the Bode gain K_B must be increased by a factor of 2 to satisfy the steady state specification. But the Bode gain of a lag network is

$$\lim_{s\to 0} P_{Lag} = \lim_{s\to 0}\frac{1+s/b}{1+s/a} = 1$$

Therefore the compensation required in this problem has the form $2(1+s/a)/(1+s/b)$ where the twofold gain factor increase is supplied by an amplifier and a and b for the lag network must be chosen to satisfy the requirements on M_p and ϕ_{PM}. The gain-compensated function is shown as $G_1(j\omega) = 2G(j\omega)$ in Fig. 18-18; $G_1(j\omega)$ must be shifted downward by 7 to 10 db in the region $0.7 \leq \omega \leq 2.0$, with no substantial increase in phase lag, to meet the transient specifications.

Referring to Fig. 18-7, we see that, for $b/a = 3$, we can obtain a maximum attenuation of 9.5 db. For $a = 0.1$, the phase lag is $-15°$ at $\omega = 0.7$ ($\omega/a = 7$) and $-6°$ at $\omega = 2.0$ ($\omega/a = 20$), that is, the phase lag is relatively small in the frequency region of interest. The db magnitude-phase angle plot for

$$G_4(j\omega) = \frac{2(1+j\omega/0.3)}{1+j\omega/0.1}G(j\omega)$$

is also shown in Fig. 18-18, with $M_p \cong 5$ db and $\phi_{PM} = 32°$; hence this system does not meet the specifications. To decrease the phase lag introduced in the frequency region $0.7 \leq \omega \leq 2.0$, we change a to 0.05 and b to 0.15. The phase lag is now 9° at $\omega = 0.7$ ($\omega/a = 14$). The db magnitude-phase angle plot for

$$G_5(j\omega) = \frac{2(1+j\omega/0.15)}{1+j\omega/0.05}G(j\omega)$$

is shown in Fig. 18-18. We see that $M_p \cong 3.5$ db and $\phi_{PM} = 41°$. Thus the specifications are satisfied. The desired compensation is given by

$$2P_{Lag} = \frac{2(1+s/0.15)}{1+s/0.05}$$

18.9. Solve Problem 18.7 using *lag-lead* plus gain compensation. In addition to the previous specifications, we require that the resonant frequency ω_p of the compensated system be approximately the same as that for the uncompensated system.

In Problems 18.7 and 18.8 we found that the Bode gain K_B must be increased by a factor of 2 to satisfy the steady state specification. The frequency response function of the lag-lead plus gain compensation is therefore given by

$$2P_{LL}(j\omega) = \frac{2(1+j\omega/a_1)(1+j\omega/b_2)}{(1+j\omega/b_1)(1+j\omega/a_2)}$$

We must now choose a_1, b_1, b_2, and a_2 to satisfy the requirements on M_p, ϕ_{PM} and ω_p. Referring to Fig. 18-17, we see that the resonant frequency for the uncompensated system is about 1.1 rad/sec. The db magnitude-phase angle plot of $G_1(j\omega) = 2G(j\omega)$ shown in Fig. 18-19 indicates that, if the $G_1(j\omega)$ curve is attenuated by 6.5 db and 10° of phase lead is added at a frequency of $\omega = 1.0$ rad/sec, then the resulting curve will be tangent to the $M_p = 2$ db curve at about 1 rad/sec. Referring to Fig. 18-10, if we let $b_1/a_1 = b_2/a_2 = 3$, $a_1 = 6a_2$, and $\omega/a_2 = 6.0$ for $\omega = 1$, we obtain the desired attenuation and phase lead. Solving for the remaining parameters, we get $a_2 = 1/6 = 0.167$, $b_2 = 3a_2 = 0.50$, $a_1 = 6a_2 = 1.0$, $b_1 = 3a_1 = 3.0$. The db magnitude-phase angle plot for the resulting open-loop frequency response function

$$G_6(j\omega) = \frac{2(1+j\omega)(1+j\omega/0.5)}{(1+j\omega/3)(1+j\omega/0.167)}G(j\omega)$$

is shown in Fig. 18-19, where $M_p \cong 3.5$ db, $\phi_{PM} = 44°$, and $\omega_p \cong 1.0$ rad/sec. These values approximately satisfy the specifications.

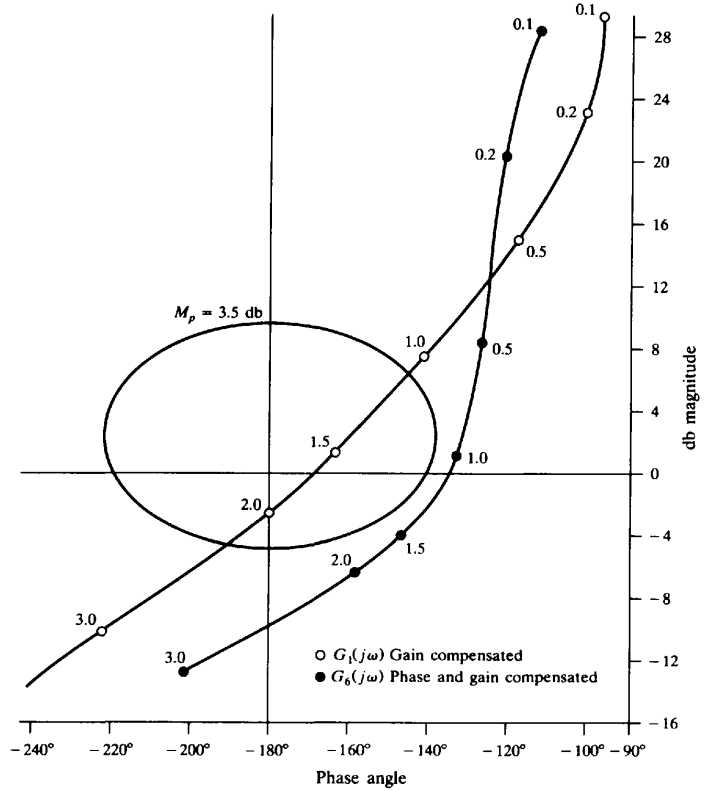


Fig. 18-19

18.10. Design compensation for the discrete-time system with open-loop transfer function

$$GH(z) = \frac{K(z+1)^3}{(z-1)\left(z+\frac{2}{3}\right)^2}$$

such that the following performance specifications are satisfied:

1. gain margin ≥ 6 db
2. phase margin $\phi_{PM} \geq 45°$
3. gain crossover frequency ω_1 such that $\omega_1 T \leq 1.6$ rad
4. velocity constant $K_v \geq 10$

The Nichols chart plot of GH shown in Fig. 18-20 indicates that $\omega_1 T = 1.6$ rad for $K = -3$ db. The gain and phase margins are met if $K < 4.7$ db; but the steady state specification requires that $K > 10.8$ db (gain factor of 3.47). Substituting $z = (1+w)/(1-w)$, we transform the open-loop transfer function from the z-domain to the w-domain, thus forming

$$GH'(w) = \frac{36}{25}\frac{K}{w(1+w/5)^2}$$

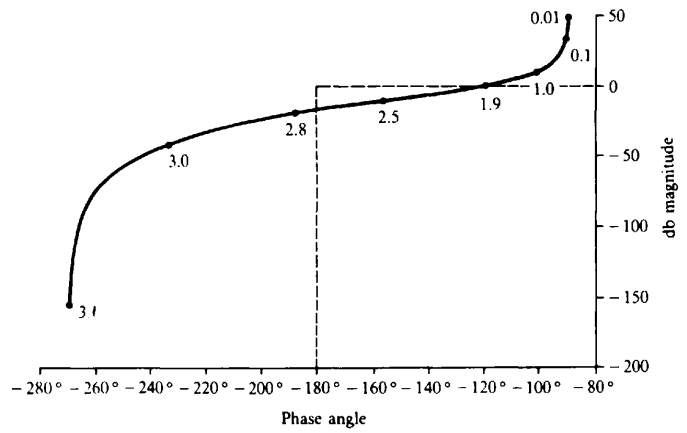


Fig. 18-20

In the w-domain the gain crossover frequency specification becomes

$$\omega_{w1} = \tan\left(\frac{\omega_1 T}{2}\right) = 1.02 \text{ rad/sec}$$

A low-frequency cascade lag compensator with $b/a = 3.5$ can be used to increase K_v to 10, while maintaining the gain crossover frequency ω_1 and the gain and phase margins at their previous values. A lag compensator with $b = 0.35$ and $a = 0.1$ satisfies the requirements.

The lag compensator in the w-plane is

$$G_1(w) = \frac{3.5(1+w/0.35)}{1+w/0.1}$$

This is transformed back into the z-domain by substituting $w = (z-1)/z+1)$, thus forming

$$G_1(z) = 1.2273\left(\frac{z-0.4815}{z-0.8182}\right)$$

The db magnitude-phase angle plot for the compensated discrete-time system is shown in Fig. 18-21.

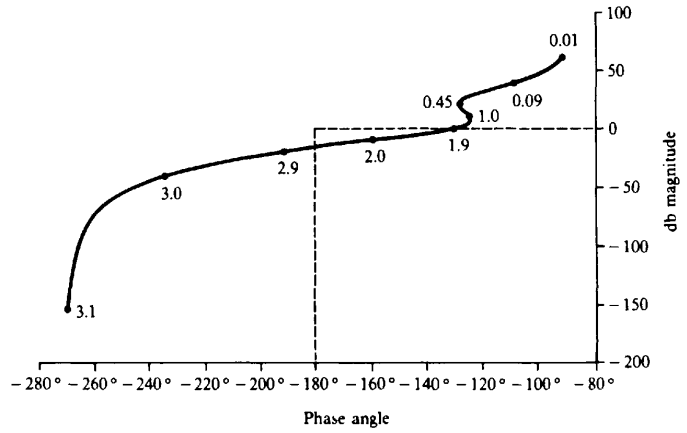


Fig. 18-21

Supplementary Problems

18.11. Find a value of K_B for which the system whose open-loop transfer function is

$$GH = \frac{K_B}{s(1 + s/200)(1 + s/250)}$$

has a resonant peak M_p of 1.4 db. *Ans.* $K_B = 119.4$.

18.12. For the system of Problem 18.11, find gain plus lag compensation such that $M_p \le 1.7$, $\phi_{PM} \ge 35°$, and $K_v \ge 50$.

18.13. For the system of Problem 18.11, find gain plus lead compensation such that $M_p \le 1.7$, $\phi_{PM} \ge 50°$, and $K_v \ge 50$.

18.14. For the system of Problem 18.11, find gain plus lag-lead compensation such that $M_p \le 1.5$, $\phi_{PM} \ge 40°$, and $K_v \ge 100$.

18.15. Find gain plus lag compensation for the system whose open-loop transfer function is

$$GH = \frac{K_B}{s(1 + s/10)(1 + s/5)}$$

such that $K_v = 30$ and $\phi_{PM} \ge 40°$.

18.16. For the system of Problem 18.15, find gain plus lead compensation such that $K_v \ge 30$ and $\phi_{PM} \ge 45°$. *Hint.* Cadcade two lead compensation networks.

18.17. Find gain plus lead compensation for the system whose open-loop transfer function is

$$GH = \frac{K_B}{s(1 + s/2)}$$

such that $K_v = 20$ and $\phi_{PM} = 45°$.

Chapter 19

Introduction to Nonlinear Control Systems

19.1 INTRODUCTION

We have thus far confined the discussion to systems describable by linear time-invariant ordinary differential or difference equation models or their transfer functions, excited by Laplace or z-transformable input functions. The techniques developed for studying these systems are relatively straightforward and usually lead to practical control system designs. While it is probably true that no physical system is *exactly* linear and time-invariant, such models are often adequate approximations and, as a result, the linear system methods developed in this book have broad application. There are many situations, however, for which linear representations are inappropriate and *nonlinear* models are required.

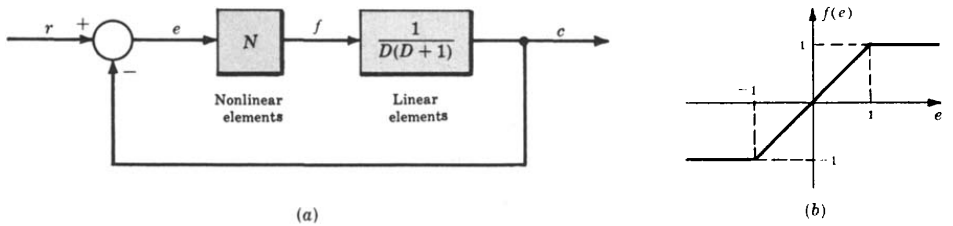
Theories and methods for analysis and design of nonlinear control systems constitute a large body of knowledge, some of it quite complex. The purpose of this chapter is to introduce some of the prevailing classical techniques, utilizing mathematics at about the same level as in earlier chapters.

Linear systems are defined in Definition 3.21. Any system that does not satisfy this definition is nonlinear. The major difficulty with nonlinear systems, especially those described by nonlinear ordinary differential or difference equations, is that analytical or closed-form solutions are available only for very few special cases, and these are typically not of practical interest in control system analysis or design. Furthermore, unlike linear systems, for which free and forced responses can be determined separately and the results superimposed to obtain the total response, free and forced responses of nonlinear systems normally *interact* and cannot be studied separately, and superposition does not generally hold for inputs or initial conditions.

In general, the characteristic responses and stability of nonlinear systems depend qualitatively as well as quantitatively on initial condition values, and the magnitude, shape, and form of system inputs. On the other hand, time-domain solutions to nonlinear system equations usually can be obtained, for *specified* inputs, parameters, and initial conditions, by computer simulation techniques. Algorithms and software for simulation, a special topic outside the scope of this book, are widely available and therefore are not developed further here. Instead, we focus on several analytical methods for studying nonlinear control systems.

Nonlinear control system problems arise when the structure or fixed elements of the system are inherently nonlinear, and/or nonlinear compensation is introduced into the system for the purpose of improving its behavior. In either case, stability properties are a central issue.

EXAMPLE 19.1. Fig. 19-1(a) is a block diagram of a nonlinear feedback system containing two blocks. The *linear* block is represented by the transfer function $G_2 = 1/D(D + 1)$, where $D \equiv d/dt$ is the *differential* operator. D is used instead of s in this linear transfer function because the Laplace transform and its inverse are generally not strictly applicable for nonlinear analysis of systems with both linear and nonlinear elements. Alternatively, when using the describing function method (Section 19.5), an approximate frequency response technique, we



(a) (b)

Fig. 19-1

usually write

$$G_2(j\omega) = \frac{1}{j\omega(j\omega + 1)}$$

The *nonlinear block N* has the transfer characteristic $f(e)$ defined in Figure 19-1(b). Such nonlinearities are called (piecewise-linear) **saturation** functions, described further in the next section.

EXAMPLE 19.2. If the earth is assumed spherical and all external forces other than gravity are negligible, then the motion of an earth satellite lies in a plane called the *orbit plane*. This motion is defined by the following set of nonlinear differential equations (see Problem 3.3):

$$r\frac{d^2\theta}{dt^2} + 2\frac{dr}{dt}\frac{d\theta}{dt} = 0 \qquad \text{(transverse force equation)}$$

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = -\frac{k^2}{pr^2} \qquad \text{(radial force equation)}$$

The satellite, together with any controller designed to modify its motion, constitutes a nonlinear control system.

Several popular methods for nonlinear analysis are summarized below.

19.2 LINEARIZED AND PIECEWISE-LINEARIZED APPROXIMATIONS OF NONLINEAR SYSTEMS

Nonlinear terms in differential or difference equations can sometimes be approximated by linear terms or zero-order (constant) terms, over limited ranges of the system response or system forcing function. In either case, one or more linear differential or difference equations can be obtained as approximations of the nonlinear system, valid over the same limited operating ranges.

EXAMPLE 19.3. Consider the spring-mass system of Fig. 19-2, where the spring force $f_s(x)$ is a nonlinear function of the displacement x measured from the rest position, as shown in Fig. 19-3.

The equation of motion of the mass is $M(d^2x/dt^2) + f_s(x) = 0$. However, if the absolute magnitude of the displacement does not exceed x_0, then $f_s(x) = kx$, where k is a constant. In this case, the equation of motion is a constant-coefficient linear equation given by $M(d^2x/dt^2) + kx = 0$, valid for $|x| \le x_0$.

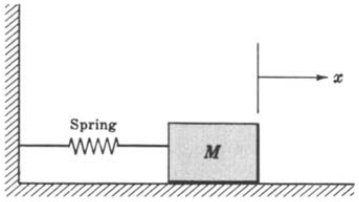


Fig. 19-2

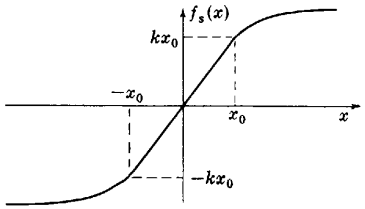


Fig. 19-3

EXAMPLE 19.4. We again consider the system of Example 19.3, but now the displacement x exceeds x_0. To treat this problem, let the spring force curve be approximated by three straight lines as shown in Fig. 19-4, a *piecewise-linear* approximation of $f_s(x)$.

The *system* is then approximated by a piecewise-linear system; that is, the system is described by the linear equation $M(d^2x/dt^2) + kx = 0$ when $|x| \le x_1$, and by the equations $M(d^2x/dt^2) \pm F_1 = 0$ when $|x| > x_1$. The + sign is used if $x > x_1$ and the − sign if $x < -x_1$.

Nonlinear terms in a system equation are sometimes known in a form that can be easily expanded in a series, for example, a Taylor or a Maclaurin series. In this manner, a nonlinear term can be approximated by the first few terms of the series, excluding terms higher than first degree.

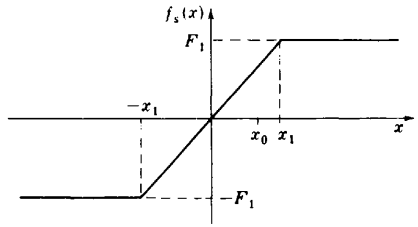


Fig. 19-4

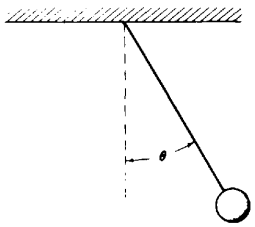


Fig. 19-5

EXAMPLE 19.5. Consider the nonlinear equation describing the motion of a pendulum (see Fig. 19-5):

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\sin\theta = 0$$

where l is the length of the pendulum bob and g is the acceleration of gravity. If small motions of the pendulum about the "operating point" $\theta = 0$ are of interest, then the equation of motion can be linearized about this operating point. This is done by forming a Taylor series expansion of the nonlinear term $(g/l)\sin\theta$ about the point $\theta = 0$ and retaining only the first degree terms. The nonlinear equation is

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\sin\theta = \frac{d^2\theta}{dt^2} + \frac{g}{l}\sum_{k=0}^{\infty}\frac{\theta^k}{k!}\left(\frac{d^k}{d\theta^k}(\sin\theta)\bigg|_{\theta=0}\right)$$

$$= \frac{d^2\theta}{dt^2} + \frac{g}{l}\left[\theta - \frac{\theta^3}{3!} + \cdots\right] = 0$$

The linear equation is $d^2\theta/dt^2 + (g/l)\theta = 0$, valid for small variations in θ.

It is instructive to express the linearization process more formally for Taylor series applications, to better establish its applicability and limitations.

Taylor Series

The infinite series expansion of a general nonlinear function $f(x)$ can be quite useful in nonlinear systems analysis. The function $f(x)$ can be written as the following infinite series, expanded about the point $\bar{x}$:

$$f(x) = f(\bar{x}) + \frac{df}{dx}\bigg|_{x=\bar{x}}(x-\bar{x}) + \frac{1}{2!}\frac{d^2f}{dx^2}\bigg|_{x=\bar{x}}(x-\bar{x})^2 + \cdots$$

$$= \sum_{k=0}^{\infty}\frac{(x-\bar{x})^k}{k!}\frac{d^kf}{dx^k}\bigg|_{x=\bar{x}} \qquad (19.1)$$

where $(d^kf/dx^k)|_{x=\bar{x}}$ is the value of the kth derivative of f with respect to x evaluated at the point $x = \bar{x}$. Clearly, this expansion exists (is feasible) only if all the required derivatives exist.

If the sum of the terms of Equation (*19.1*) second-degree and higher-degree in $(x - \bar{x})$ are negligible compared with the sum of the first two terms, then we can write

$$f(x) \cong f(\bar{x}) + \frac{df}{dx}\bigg|_{x=\bar{x}}(x-\bar{x}) \qquad (19.2)$$

This approximation usually works if x is "close enough" to $\bar{x}$, or, equivalently, if $x - \bar{x}$ is "small enough," in which case higher-degree terms are *relatively* small.

Equation (*19.2*) can be rewritten as

$$f(x) - f(\bar{x}) \cong \left.\frac{df}{dx}\right|_{x=\bar{x}} (x - \bar{x}) \tag{19.3}$$

Then if we define

$$\Delta x \equiv x - \bar{x} \tag{19.4}$$

$$\Delta f \equiv f(x) - f(\bar{x}) \tag{19.5}$$

Equation (*19.3*) becomes

$$\Delta f \cong \left.\frac{df}{dx}\right|_{x=\bar{x}} \Delta x \tag{19.6}$$

If $x = x(t)$ is a function of time t, or any other independent variable, then in most applications t can be treated as a fixed parameter when performing the linearization computations above, and $\Delta x = \Delta x(t) \equiv x(t) - \bar{x}(t)$, etc.

EXAMPLE 19.6. Suppose $y(t) = f[u(t)]$ represents a nonlinear system with input $u(t)$ and output $y(t)$, where $t \geq t_0$ for some t_0, and df/du exists for all u. If the normal operating conditions for this system are defined by the input $u = \bar{u}$ and output $y = \bar{y}$, then small changes $\Delta y(t) = y(t) - \bar{y}(t)$ in output operation in response to small changes in the input $\Delta u(t) = u(t) - \bar{u}(t)$ can be expressed by the approximate linear relation

$$\Delta y(t) \cong \left.\frac{df}{du}\right|_{u=\bar{u}(t)} \Delta u(t) \tag{19.7}$$

for $t \geq t_0$.

Taylor Series for Vector Processes

Equations (*19.1*) through (*19.7*) are readily generalized for nonlinear m-vector functions of n-vector arguments, $\mathbf{f}(\mathbf{x})$, where

$$\mathbf{f} \equiv \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} \qquad \mathbf{x} \equiv \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

and m and n are arbitrary. In this case, $\Delta\mathbf{x} \equiv \mathbf{x} - \bar{\mathbf{x}}$, $\Delta\mathbf{f} \equiv \mathbf{f}(\mathbf{x}) - \mathbf{f}(\bar{\mathbf{x}})$, and Equation (*19.6*) becomes

$$\Delta\mathbf{f} \cong \left.\frac{d\mathbf{f}}{d\mathbf{x}}\right|_{x=\bar{x}} \Delta\mathbf{x} \tag{19.8}$$

where $d\mathbf{f}/d\mathbf{x}$ is a matrix defined as

$$\frac{d\mathbf{f}}{d\mathbf{x}} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{19.9}$$

EXAMPLE 19.7. For $m = 1$ and $n = 2$, Equation (*19.9*) reduces to

$$\frac{df}{d\mathbf{x}} = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} \end{bmatrix}$$

and Equation (*19.8*) is

$$\Delta f \cong \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \frac{\partial f}{\partial x_1}\Delta x_1 + \frac{\partial f}{\partial x_2}\Delta x_2 \tag{19.10}$$

Equation (*19.10*) represents the common case where a nonlinear scalar function f of two variables, say $x_1 \equiv x$ and $x_2 \equiv y$, are linearized about a point $\{\bar{x}, \bar{y}\}$ in the plane.

Linearization of Nonlinear Differential Equations

We follow the same procedure to linearize differential equations as we did above in linearizing functions $\mathbf{f}(x)$. Consider a **nonlinear differential system** written in state variable form:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t)] \tag{19.11}$$

where the vector of n state variables $\mathbf{x}(t)$ and the r-input vector $\mathbf{u}(t)$ are defined as in Chapter 3, Equations (*3.24*) and (*3.25*), and $t \geq t_0$. In Equation (*19.11*), $\mathbf{f}$ is an n-vector of nonlinear functions of $\mathbf{x}(t)$ and $\mathbf{u}(t)$.

Similarly, **nonlinear output equations** may be written in vector form:

$$\mathbf{y}(t) = \mathbf{g}[\mathbf{x}(t)] \tag{19.12}$$

where $\mathbf{y}(t)$ is an m-vector of outputs and $\mathbf{g}$ is an m-vector of nonlinear functions of $\mathbf{x}(t)$.

EXAMPLE 19.8. One example of a nonlinear SISO differential system of the form of Equations (*19.11*) and (*19.12*) is

$$\frac{dx_1}{dt} = f_1(\mathbf{x}, u) = c_1 u x_2 - c_2 x_1^2$$

$$\frac{dx_2}{dt} = f_2(\mathbf{x}, u) = \frac{c_3 x_1}{c_4 + x_1}$$

$$y = g(\mathbf{x}) = c_5 x_1^2$$

The *linearized versions* of Equations (*19.11*) and (*19.12*) are given by

$$\frac{d(\Delta\mathbf{x})}{dt} \cong \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\substack{\mathbf{x}=\bar{\mathbf{x}}(t) \\ \mathbf{u}=\bar{\mathbf{u}}(t)}} \Delta\mathbf{x} + \left.\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\right|_{\substack{\mathbf{x}=\bar{\mathbf{x}}(t) \\ \mathbf{u}=\bar{\mathbf{u}}(t)}} \Delta\mathbf{u} \tag{19.13}$$

$$\Delta y(t) \cong \left.\frac{\partial \mathbf{g}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\bar{\mathbf{x}}(t)} \Delta\mathbf{x} \tag{19.14}$$

where the partial derivative matrices in these equations are defined as in Equations (*19.9*) and (*19.10*), each evaluated at the "point" $\{\bar{\mathbf{x}}, \bar{\mathbf{u}}\}$. The pair $\bar{\mathbf{x}} \equiv \bar{\mathbf{x}}(t)$ and $\bar{\mathbf{u}} \equiv \bar{\mathbf{u}}(t)$ are actually functions of time, but they are treated like "points" in the indicated computations.

Linearized equations (*19.13*) and (*19.14*) are usually interpreted as follows. If the input is perturbed or deviates from an "operating point" $\bar{\mathbf{u}}(t)$ by a small enough amount $\Delta\mathbf{u}(t)$, generating small enough perturbations $\Delta\mathbf{x}(t)$ in the state and small enough perturbations in the output $\Delta\mathbf{y}(t)$ about their operating points, then the *linear* equations (*19.13*) and (*19.14*) are reasonable approximation equations for the perturbed states $\Delta\mathbf{x}(t)$ and perturbed outputs $\Delta\mathbf{y}(t)$.

Linearized equations (*19.13*) and (*19.14*) are often called the (*small*) **perturbation equations** for the nonlinear differential system. They are *linear* in $\Delta\mathbf{x}$ and ($\Delta\mathbf{u}$), because the coefficient matrices:

$$\left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\substack{\mathbf{x}=\bar{\mathbf{x}}(t) \\ \mathbf{u}=\bar{\mathbf{u}}(t)}} \qquad \left.\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\right|_{\substack{\mathbf{x}=\bar{\mathbf{x}}(t) \\ \mathbf{u}=\bar{\mathbf{u}}(t)}} \qquad \left.\frac{\partial \mathbf{g}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\bar{\mathbf{x}}(t)}$$

having been evaluated at $\bar{\mathbf{x}}(t)$ and/or $\bar{\mathbf{u}}(t)$, are *not* functions of $\Delta\mathbf{x}(t)$ [or $\Delta\mathbf{u}(t)$].

Linearized equations (*19.13*) and (*19.14*) are also *time-invariant* if $\bar{\mathbf{u}}(t) = \bar{\mathbf{u}} =$ constant and $\bar{\mathbf{x}}(t) = \bar{\mathbf{x}} =$ constant. In this case, all of the methods developed in this book for time-invariant ordinary differential systems can be applied. Nevertheless, the results must be interpreted judiciously because, again, the linearized model is an approximation, valid only for "small enough" perturbations about an operating point and, generally speaking, "small enough" perturbations are not always easy to ascertain.

EXAMPLE 19.9. The linearized (perturbation) equations for the system given in Example 19.8 are determined as follows from Equations (*19.13*) and (*19.14*). For convenience, we first define

$$\left.\frac{\partial f}{\partial x}\right|_{\substack{x=\bar{x}(t)\\u=\bar{u}(t)}} \equiv \frac{\partial \bar{f}}{\partial x}$$

etc., to simplify the notation. Then

$$\frac{d(\Delta x_1)}{dt} = \frac{\partial \bar{f}_1}{\partial x_1}\Delta x_1 + \frac{\partial \bar{f}_1}{\partial x_2}\Delta x_2 + \frac{\partial \bar{f}_1}{\partial u}\Delta u = -2c_2\bar{x}_1\Delta x_1 + c_1\bar{u}\Delta x_2 + c_1\bar{x}_2\Delta u$$

Similarly,

$$\frac{d(\Delta x_2)}{dt} \cong \frac{\partial \bar{f}_2}{\partial x_1}\Delta x_1 + \frac{\partial \bar{f}_2}{\partial x_2}\Delta x_2 + \frac{\partial \bar{f}_2}{\partial u}\Delta u$$
$$= \frac{c_3c_4}{(c_4+\bar{x}_1)^2}\Delta x_1 + 0 + 0 = \frac{c_3c_4\,\Delta x_1}{(c_4+\bar{x}_1)^2}$$

and the output perturbation equation is

$$\Delta y \cong \frac{\partial \bar{g}}{\partial x_1}\Delta x_1 + \frac{\partial \bar{g}}{\partial x_2}\Delta x_2 = 2c_5\bar{x}_1\Delta x_1$$

Linearization of Nonlinear Discrete-Time Equations

The Taylor series linearization procedure can be applied to many discrete-time system problems, but sufficient care must be taken to justify the existence of the series. The application is often justified if the discrete-time equations represent reasonably well-behaved nonlinear processes, such as discrete-time representations of continuous systems with state variables expressed only at discrete-time instants.

EXAMPLE 19.10. The time-invariant discrete-time system represented by the nonlinear difference equation $x(k+1) = ax^2(k)$, with $a < 0$ and $x(0) \neq 0$, is easily linearized, because the nonlinear term $ax^2(k)$ is a smooth function of x. We have

$$x(k+1) = ax^2(k) \equiv f(x)$$
$$\Delta f = f(x) - f(\bar{x})$$
$$\left.\frac{\partial f}{\partial x}\right|_{x=\bar{x}} = 2a\bar{x}$$
$$x(k) = \bar{x}(k) + \Delta x(k)$$
$$\bar{x}(k+1) = a\bar{x}^2(k)$$

Substitution of these equations into Equation (*19.6*) and rearranging terms yields

$$\Delta x(k+1) \cong 2a\bar{x}(k)\,\Delta x(k)$$

which is linear in Δx, but time-varying in general.

19.3 PHASE PLANE METHODS

In Sections 3.15 and 4.6, the state variable form of linear differential equations was introduced and shown to be a useful tool for analysis of linear systems. In Section 19.2, this representation was applied to nonlinear systems via the concept of linearization. In this section, **phase plane** methods are developed for analyzing nonlinear differential equations in state variable form, without the need for linearization.

A second-order differential equation of the form:

$$\frac{d^2x}{dt^2} = f\left(x, \frac{dx}{dt}\right) \tag{19.15}$$

can be rewritten as a pair of first-order differential equations, as in Section 3.15, by making the change of variables $x = x_1$ and $dx/dt = x_2$, yielding

$$\frac{dx_1}{dt} = x_2 \tag{19.16}$$

$$\frac{dx_2}{dt} = f(x_1, x_2) \tag{19.17}$$

The two-tuple, or pair of state variables (x_1, x_2), may be considered as a point in the plane. Since x_1 and x_2 are functions of time, then as t increases, $(x_1(t), (x_2(t))$ describes a *path* or *trajectory* in the plane. This plane is called the **phase plane**, and the trajectory is a parametric plot of x_2 versus x_1, parametrized by t.

If we eliminate time as the independent variable in Equations (*19.16*) and (*19.17*), we obtain the first-order differential equation

$$\frac{dx_1}{dx_2} = \frac{x_2}{f(x_1, x_2)} \tag{19.18}$$

Solution of Equation (*19.18*) for x_1 as a function of x_2 (or vice versa) defines a trajectory in the phase plane. By solving this equation for various initial conditions on x_1 and x_2 and examining the resulting phase plane trajectories, we can determine the behavior of the second-order system.

EXAMPLE 19.11. The differential equation

$$\frac{d^2x}{dt^2} + \left(\frac{dx}{dt}\right)^2 = 0$$

with the initial conditions $x(0) = 0$ and $(dx/dt)|_{t=0} = 1$, can be replaced by the two first-order equations

$$\frac{dx_1}{dt} = x_2 \qquad x_1(0) = 0$$
$$\frac{dx_2}{dt} = -x_2^2 \qquad x_2(0) = 1$$

where $x \equiv x_1$ and $dx/dt \equiv x_2$. Eliminating time as the independent variable, we obtain

$$\frac{dx_1}{dx_2} = -\frac{x_2}{x_2^2} = -\frac{1}{x_2} \quad \text{or} \quad dx_1 = -\frac{dx_2}{x_2}$$

Integration of this equation for the given initial conditions yields

$$\int_{x_1(0)=0}^{x_1} dx_1' = x_1 = -\int_{x_2(0)=1}^{x_2} \frac{dx_2'}{x_2'} = -\ln x_2 \quad \text{or} \quad x_2 = e^{-x_1}$$

The phase plane trajectory defined by this equation is plotted in Fig. 19-6. Its direction in the phase plane is

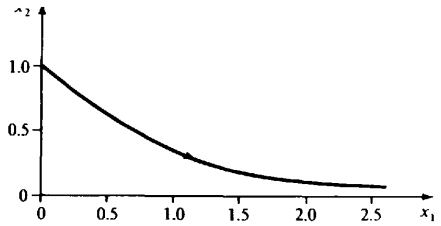


Fig. 19-6

determined by noting that $dx_2/dt = -x_2^2 < 0$ for all $x_2 \neq 0$. Therefore x_2 always decreases and we obtain the trajectory shown.

On-Off Control Systems

A particularly useful application of phase plane methods is designing *on-off controllers* (Definition 2.25), for the special class of feedback control systems with linear continuous-time second-order plants, as in Fig. 19-7 and Equation (*19.19*).

$$\frac{d^2c}{dt^2} + a\frac{dc}{dt} = u \qquad a \geq 0 \tag{19.19}$$

The initial conditions $c(0)$ and $(dc/dt)|_{t=0}$ for Equation (*19.19*) are arbitrary. The on-off controller with input $e = r - c$ generates the control signal u which attains only two values, $u = \pm 1$.

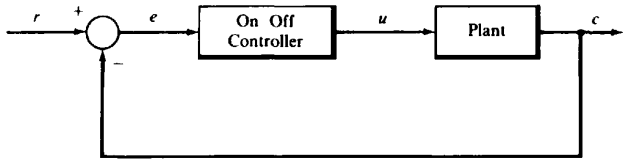


Fig. 19-7

On-Off Controller Design Specifications

If the reference input r is a unit step function applied at time zero, typical design specifications for the system of Fig. 19-7 are the following. The control input u to the plant must drive the plant output $c(t)$ to $c(t') = 1$, and its derivative dc/dt to $(dc/dt)|_{t=t'} = 0$, simultaneously, *and* in the minimum possible time t'. The steady state error becomes zero at t' and remains zero if the control signal is turned off ($u = 0$).

Since t' is required to be minimum, this is an *optimal control problem* (see Section 20.5). It can be shown that t' is minimized only if the control signal u switches values, from $+1$ to -1 or from -1 to $+1$, at most once during the time interval $0 \leq t \leq t'$.

On-Off Controller Design

In solving this design problem, it is convenient to use the error $e = r - c$, where $r = \mathbf{1}(t)$, as the variable of interest, rather than the controlled output c, because $e = 0$ and $de/dt = 0$ when $c = 1$ and $dc/dt = 0$. Therefore requiring that the error e and its derivative go to zero in minimum time is equivalent to our original problem.

To solve the problem, we first generate a differential equation for e:

$$\frac{de}{dt} = \frac{d}{dt}(r - c) = -\frac{dc}{dt}$$

$$\frac{d^2e}{dt^2} = -\frac{d^2c}{dt^2} = a\frac{dc}{dt} - u = -a\frac{de}{dt} - u \tag{19.20}$$

with initial conditions $e(0) = 1 - c(0)$ and $(de/dt)|_{t=0} = -(dc/dt)|_{t=0}$. Then we replace Equation (*19.20*) with two first-order differential equations, by letting $e \equiv x_1$ and $de/dt \equiv x_2$:

$$\frac{dx_1}{dt} = x_2 \tag{19.21}$$

$$\frac{dx_2}{dt} = -ax_2 - u \tag{19.22}$$

with initial conditions $x_1(0) = e(0) = 1 - c(0)$ and $x_2(0) = (de/dt)|_{t=0} = -(dc/dt)|_{t=0}$. Eliminating time as the independent variable, we obtain

$$\frac{dx_2}{dx_1} = -\frac{ax_2 + u}{x_2} \qquad \text{or} \qquad dx_1 = -\frac{x_2\,dx_2}{ax_2 + u} \tag{19.23}$$

This equation plus the initial conditions on $x_1(0)$ and $x_2(0)$ define a trajectory in the phase plane.

Since the control signal u switches ($+1$ to -1 or -1 to $+1$) no more than once, we can separate the trajectory into two parts, the first prior to the switching time and the second after switching. We consider the second part first, as it terminates at the origin of the phase plane, $x_1 = x_2 = 0$. We set $u = \pm 1$ in Equation (*19.23*) and then integrate between a general set of initial conditions $x_1(t)$ and $x_2(t)$ and the terminal conditions $x_1 = x_2 = 0$. To perform the integration, we consider four different sets of initial conditions, each corresponding to one of the quadrants of the phase plane.

In the first quadrant, $x_1 > 0$ and $x_2 > 0$. Note that $dx_1/dt = x_2 > 0$. Thus x_1 increases when x_2 is in the first quadrant, and when x_2 goes to zero, x_1 cannot be zero. Therefore trajectories which start in the first quadrant cannot terminate at the origin of the phase plane if u does not switch.

An identical argument holds when the initial conditions are in the third quadrant, that is, if $x_1 < 0$ and $x_2 < 0$, the trajectory cannot terminate at the origin if u does not switch.

In the second quadrant, $x_1 < 0$ and $x_2 > 0$. Since $dx_1/dt = x_2 > 0$, x_1 will increase as long as $x_2 > 0$. Since $a > 0$, then $-ax_2 < 0$ and thus $dx_2/dt < 0$ for $u = +1$ whenever $x_2 > 0$. Integration of Equation (*19.23*) with $u = +1$, initial conditions in the second quadrant, and terminal conditions $x_1 = x_2 = 0$, yields

$$\int_{x_1(t)}^{0} dx_1 = -x_1(t) = -\int_{x_2(t)}^{0} \frac{x_2\,dx_2}{ax_2 + 1}$$

or

$$x_1(t) = \frac{1}{a^2}\left[ax_2 + 1 - \ln(ax_2 + 1)\right]\Big|_{x_2(t)}^{0} = -\frac{x_2(t)}{a} + \frac{1}{a^2}\ln\left[ax_2(t) + 1\right] \tag{19.24}$$

where $x_1(t) \leq 0$, $x_2(t) \geq 0$. This equation defines a curve in the second quadrant of the phase plane such that, for any point on this curve, the trajectory terminates at the origin if $u = +1$. That is, the control signal $u = +1$ drives x_1 and x_2 to zero simultaneously.

By an identical argument, there exists a curve in the fourth quadrant defined by

$$x_1(t) = -\frac{x_2(t)}{a} - \frac{1}{a^2}\ln\left[-ax_2(t) + 1\right] \tag{19.25}$$

where $x_1(t) \geq 0$, $x_2(t) \leq 0$ such that for any $(x_1(t), x_2(t))$ on this curve the control signal $u = -1$ drives x_1 and x_2 to zero simultaneously.

The curves defined by Equations (*19.24*) and (*19.25*) join at $x_1 = x_2 = 0$ and together define the **switching curve** for the on-off controller. The switching curve divides the entire phase plane into two regions, as indicated in Fig. 19-8. The part of any trajectory after switching always starts on this curve, moves along the curve, and terminates at $x_1 = x_2 = 0$.

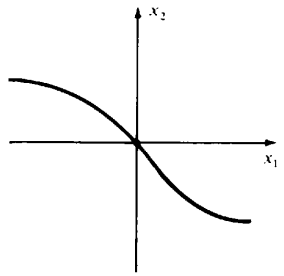


Fig. 19-8

Now we consider the part of the trajectory prior to switching. First, we explore a monotone property of the switching curve. In the second quadrant, where $u = +1$, $x_2 > 0$, and the slope of the curve is negative:

$$\frac{dx_2}{dx_1} = -\left(a + \frac{1}{x_2}\right) < 0$$

In the fourth quadrant, where $u = -1$, $x_2 < 0$, and

$$\frac{dx_2}{dx_1} = -\left(a - \frac{1}{x_2}\right) < 0$$

Therefore the slope of the entire switching curve is negative for all (x_1, x_2) on the curve, that is, the switching curve is *monotone decreasing*. Thus, corresponding to any specific value of x_1, there is one and only one corresponding value of x_2. Because of the monotone property of the switching curve, the region above the switching curve is the same as the region to the right of the switching curve, that is, it consists of the set of points (x_1, x_2) such that

$$x_1 > -\frac{x_2}{a} + \frac{1}{a^2}\ln(ax_2 + 1) \qquad (19.26)$$

when $x_2 \geq 0$ and

$$x_1 > -\frac{x_2}{a} - \frac{1}{a^2}\ln(-ax_2 + 1) \qquad (19.27)$$

when $x_2 \leq 0$.

We consider the part of the trajectory prior to switching, when the conditions $(x_1(0), x_2(0))$ lie above the switching curve. For this case, $u = +1$ and the first part of the trajectory is obtained by integration of Equation (*19.23*) with $u = +1$ between the initial conditions $(x_1(0), x_2(0))$ and an arbitrary pair of points $(x_1(t), x_2(t))$ which satisfy the inequalities (*19.26*) and (*19.27*). We obtain the trajectory by integrating Equation (*19.23*), which yields

$$\int_{x_1(0)}^{x_1(t)} dx_1 = x_1(t) - x_1(0) = -\int_{x_2(0)}^{x_2(t)} \frac{x_2\,dx_2}{ax_2 + 1} = -\frac{1}{a^2}\left[ax_2 + 1 - \ln(ax_2 + 1)\right]\Big|_{x_2(0)}^{x_2(t)}$$

or

$$x_1(t) = x_1(0) + \frac{x_2(0)}{a} - \frac{1}{a^2}\ln[ax_2(0) + 1] - \frac{x_2(t)}{a} + \frac{1}{a^2}\ln[ax_2(t) + 1] \qquad (19.28)$$

Note that this part of the trajectory has the same shape as that in Equation (*19.24*), but that it is shifted to the right. So, when $x_2(t) = 0$, $x_1(t) = x_1(0) + (1/a)[x_2(0) - (1/a)\ln(ax_2(0) + 1)]$, which is greater than 0 because of inequality (*19.26*).

Thus, when $(x_1(0), x_2(0))$ lies above the switching curve, the on-off controller develops a control signal $u = +1$ and the resultant trajectory $(x_1(t), x_2(t))$ is defined by Equation (*19.28*). When this trajectory intersects the switching curve, that is, when $(x_1(t), x_2(t))$ satisfies Equations (*19.25*) and (*19.28*) simultaneously, the on-off controller switches the control signal to $u = -1$ and the trajectory continues along the switching curve to the origin of the phase plane.

By identical reasoning, if the initial conditions lie below the switching curve, that is,

$$x_1(0) < -\frac{x_2(0)}{a} + \frac{1}{a^2}\ln[ax_2(0) + 1]$$

when $x_2(0) \geq 0$, or

$$x_1(0) < -\frac{x_2(0)}{a} - \frac{1}{a^2}\ln[-ax_2(0) + 1]$$

when $x_2(0) \leq 0$, then the on-off controller develops a control signal $u = -1$ and the trajectory

$(x_1(t), x_2(t))$ satisfies

$$x_1(t) = x_1(0) + \frac{x_2(0)}{a} + \frac{1}{a^2}\ln[-ax_2(0) + 1] - \frac{x_2(t)}{a} - \frac{1}{a^2}\ln[-ax_2(t) + 1] \qquad (19.29)$$

When this trajectory intersects the switching curve, that is, when $(x_1(t), x_2(t))$ satisfies Equations (*19.24*) and (*19.29*) simultaneously, the on-off controller switches the control signal to $u = +1$ and the trajectory moves along the switching curve in the second quadrant and terminates at the origin of the phase plane.

Recalling that $x_1 \equiv e$ and $x_2 \equiv \dot{e}$, the switching logic of the on-off controller is as follows:

(*a*) When $\dot{e} > 0$ and $e + \frac{\dot{e}}{a} - \frac{1}{a^2}\ln(a\dot{e} + 1) > 0$, then $u = +1$

(*b*) When $\dot{e} < 0$ and $e + \frac{\dot{e}}{a} + \frac{1}{a^2}\ln(-a\dot{e} + 1) > 0$, then $u = +1$

(*c*) When $\dot{e} > 0$ and $e + \frac{\dot{e}}{a} - \frac{1}{a^2}\ln(a\dot{e} + 1) < 0$, then $u = -1$

(*d*) When $\dot{e} < 0$ and $e + \frac{\dot{e}}{a} + \frac{1}{a^2}\ln(-a\dot{e} + 1) < 0$, then $u = -1$

EXAMPLE 19.12. For the feedback control system depicted in Fig. 19-7 and plant defined by Equation (*19.19*) with parameter $a = 1$, the switching curve is defined by

$$e = -\dot{e} + \ln(\dot{e} + 1) \qquad \text{for} \quad \dot{e} > 0$$

$$e = -\dot{e} - \ln(-\dot{e} + 1) \qquad \text{for} \quad \dot{e} < 0$$

and the switching logic for the on-off controller is given in Table 19.1.

Table 19.1

$\dot{e} > 0$	$f_1(e) = e + \dot{e} - \ln(\dot{e} + 1) > 0$	$f_2(e) = e + \dot{e} + \ln(-\dot{e} + 1) > 0$	u
No	No	No	−1
No	No	Yes	+1
No	Yes	No	−1
No	Yes	Yes	+1
Yes	No	No	−1
Yes	No	Yes	−1
Yes	Yes	No	+1
Yes	Yes	Yes	+1

Generalization

Phase plane methods apply to second-order systems. The approach has been generalized to third- and higher-order systems, but the analysis is typically much more complex. For example, to design on-off controllers in this way for third-order systems, switching curves are replaced by switching *surfaces* and the switching logic becomes far more extensive than that given in Table 19.1 for second-order systems.

19.4 LYAPUNOV'S STABILITY CRITERION

The stability criteria presented in Chapter 5 cannot be applied to nonlinear systems in general, although they may be applicable if the system is linearized, as in Section 19.2, if the perturbations $\Delta\mathbf{x}$ are small enough, and if $\bar{\mathbf{u}}(t)$ and $\bar{\mathbf{x}}(t)$ are constant, that is, if the linearized equations are time-

invariant. A more general method is provided by the Lyapunov theory, for exploring the stability of system states $\mathbf{x}(t)$ and outputs $\mathbf{y}(t)$ in the time domain, for any size perturbations $\Delta\mathbf{x}(t)$. It can be used for both linear and nonlinear systems described by sets of simultaneous first-order ordinary differential or difference equations, which we write concisely here in state variable form:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}) \tag{19.30}$$

or

$$\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k)] \tag{19.31}$$

The following stability definitions are for unforced systems, that is, for $\mathbf{u} = \mathbf{0}$, and for simplicity we write $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ or $\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k)]$.

A point $\mathbf{x}_s$ for which $\mathbf{f}(\mathbf{x}_s) = \mathbf{0}$ is called a **singular point**. A singular point $\mathbf{x}_s$ is said to be **stable** if, for any hyperspherical region S_R (e.g., a circle in two dimensions) of radius R centered at $\mathbf{x}_s$, there exists a hyperspherical region S_r of radius $r \le R$ also centered at $\mathbf{x}_s$ in which any motion $\mathbf{x}(t)$ of the system beginning in S_r remains in S_R ever after.

A singular point $\mathbf{x}_s$ is **asymptotically stable** if it is stable and all trajectories (motions) $\mathbf{x}(t)$ tend toward $\mathbf{x}_s$ as time goes to infinity.

The **Lyapunov stability criterion** states that, if the origin is a singular point, then it is stable if a **Lyapunov function** $V(\mathbf{x})$ can be found with the following properties:

(a) $V(\mathbf{x}) > 0$ for all values of $\mathbf{x} \neq \mathbf{0}$ (19.32)

(b) $dV/dt \le 0$ for all $\mathbf{x}$, for continuous systems, or $\Delta V[\mathbf{x}(k)] \equiv V[\mathbf{x}(k+1)] - V[\mathbf{x}(k)] \le 0$, for all $\mathbf{x}$, for discrete-time systems (19.33)

Furthermore, if dV/dt (or ΔV) is never zero except at the origin, the origin is *asymptotically stable*.

EXAMPLE 19.13. A nonlinear continuous system represented by

$$\frac{d^2x}{dt^2} + \frac{dx}{dt} + \left(\frac{dx}{dt}\right)^3 + x = 0$$

or, equivalently, the pair of equations

$$\frac{dx_1}{dt} = x_2 \qquad \frac{dx_2}{dt} = -x_2 - x_2^3 - x_1$$

where $x_1 \equiv x$, has a singular point at $x_1 = x_2 = 0$. The function $V = x_1^2 + x_2^2$ is positive for all x_1 and x_2, except $x_1 = x_2 = 0$ where $V = 0$. The derivative

$$\frac{dV}{dt} = 2x_1\frac{dx_1}{dt} + 2x_2\frac{dx_2}{dt} = 2x_1x_2 + 2x_2(-x_2 - x_2^3 - x_1) = -2x_2^2 - 2x_2^4$$

is never positive. Therefore the origin is stable.

EXAMPLE 19.14. The nonlinear system shown in Fig. 19-9 is represented by the differential equations [with $x_1(t) \equiv -c(t)$]:

$$\dot{x}_1 = -x_1 + x_2$$
$$\dot{x}_2 = -f(x_1 + r)$$

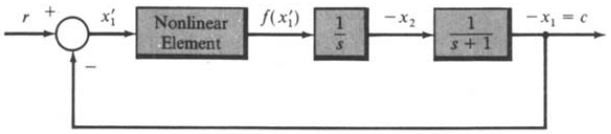


Fig. 19-9

Also, $f(0) \equiv 0$ for this particular nonlinear element. If r is constant, we can make the changes of variables $x_1' \equiv x_1 + r$, $x_2' \equiv x_2 + r$, and the state equations become

$$\dot{x}_1' = -x_1' + x_2'$$
$$\dot{x}_2' = -f(x_1')$$

The origin $x_1' = x_2' = 0$ is a singular point since $\dot{x}_1' = \dot{x}_2' = 0$ at the origin. The Lyapunov function is defined by $V \equiv 2\int_0^{x_1'} f(e)\,de + x_2'^2 > 0$ for all $x_1', x_2' \neq 0$, if $x_1' f(x_1') > 0$ for all $x_1' \neq 0$. Differentiating V,

$$\dot{V} = 2f(x_1')\dot{x}_1' + 2x_2'\dot{x}_2' = 2f(x_1')(-x_1' + x_2') - 2x_2'f(x_1') = -2x_1'f(x_1')$$

Thus, if we restrict $x_1' f(x_1') > 0$, to maintain $V > 0$, $\dot{V} \le 0$ for $x_1' \neq 0$. Therefore the system is stable for any nonlinear element satisfying the conditions

$$f(0) = 0$$
$$x_1' f(x_1') > 0 \qquad \text{for} \quad x_1' \neq 0$$

Note that this result is very general in that only the conditions above are required to assure stability.

If r is not constant, the solution for $x_1(t)$ and $x_2(t)$ corresponding to $r(t)$ is in general not constant. But, if the solution were known, the stability of the solution could be analyzed in a similar way.

EXAMPLE 19.15. For the discrete-time system

$$x_1(k+1) = x_2(k)$$
$$x_2(k+1) = -f[x_1(k)]$$

where $f(x_1)$ is the saturation nonlinearity in Fig. 19-1(b), the origin is a singular point because $x_1(k) = x_2(k) = 0$ implies $x_1(k+1) = x_2(k+1) = 0$. Let $V \equiv x_1^2 + x_2^2$, which is greater than zero for all $x_1, x_2 \neq 0$. Then

$$\begin{aligned}\Delta V &= x_1^2(k+1) + x_2^2(k+1) - x_1^2(k) - x_2^2(k)\\ &= x_2^2(k) + f^2[x_1(k)] - x_1^2(k) - x_2^2(k)\\ &= -x_1^2(k) + f^2[x_1(k)]\end{aligned}$$

Since $f^2(x_1) \le x_1^2$ for all x_1, $\Delta V \le 0$ for all x_1, x_2 and therefore the origin is stable.

Choosing Lyapunov Functions

For many problems, a convenient choice for the Lyapunov function $V(\mathbf{x})$ is the scalar quadratic form function $V(\mathbf{x}) = \mathbf{x}^T P \mathbf{x}$, where $\mathbf{x}^T$ is the transpose of the column vector $\mathbf{x}$ and P is a real symmetric matrix. To render $V > 0$, the matrix P must be *positive definite*. By Sylvester's theorem [7], P is **positive definite** if and only if all its discriminants are positive, that is,

$$P_{11} > 0$$
$$\begin{vmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{vmatrix} > 0$$
$$\vdots$$
$$\begin{vmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \vdots & \vdots \\ P_{n1} & \cdots & P_{nn} \end{vmatrix} > 0 \tag{19.34}$$

For continuous systems $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, the derivative of $V(\mathbf{x}) = \mathbf{x}^T P \mathbf{x}$ is given by

$$\dot{V}(\mathbf{x}) = \dot{\mathbf{x}}^T P \mathbf{x} + \mathbf{x}^T P \dot{\mathbf{x}} = \mathbf{f}^T(\mathbf{x}) P \mathbf{x} + \mathbf{x}^T P \mathbf{f}(\mathbf{x})$$

For discrete systems, $\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k)]$ and

$$\begin{aligned}\Delta V(k) &= V(k+1) - V(k) = \mathbf{x}^T(k+1) P \mathbf{x}(k+1) - \mathbf{x}^T(k) P \mathbf{x}(k)\\ &= \mathbf{f}^T[\mathbf{x}(k)] P \mathbf{f}[\mathbf{x}(k)] - \mathbf{x}^T(k) P \mathbf{x}(k)\end{aligned}$$

EXAMPLE 19.16. For the system represented by $\dot{\mathbf{x}} = A\mathbf{x}$ with $A = \begin{bmatrix} -2 & 1 \\ 2 & -3 \end{bmatrix}$, let $V = \mathbf{x}^T P\mathbf{x}$ with $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Then

$$\dot{V} = \mathbf{x}^T[A^T P + PA]\mathbf{x} = \mathbf{x}^T\left[\begin{bmatrix} -2 & 2 \\ 1 & -3 \end{bmatrix} + \begin{bmatrix} -2 & 1 \\ 2 & -3 \end{bmatrix}\right]\mathbf{x}$$

$$\dot{V} = \mathbf{x}^T\begin{bmatrix} -4 & 3 \\ 3 & -6 \end{bmatrix}\mathbf{x} = -\mathbf{x}^T Q\mathbf{x}$$

where $$Q = \begin{bmatrix} 4 & -3 \\ -3 & 6 \end{bmatrix}$$

Since P is positive definite, $V > 0$ for all $\mathbf{x} \neq 0$. The discriminants of Q are 4 and $(24 - 9) = 15$. Therefore Q is positive definite and $-Q$ is negative definite, which guarantees that $\dot{V} < 0$ for all $\mathbf{x} \neq 0$. The origin is therefore asymptotically stable for this system.

19.5 FREQUENCY RESPONSE METHODS

Describing Functions

Describing functions are approximate frequency response functions for the nonlinear elements of a system, which can be used to analyze the overall system using frequency response techniques developed in earlier chapters.

A describing function is developed for a nonlinear element by analyzing its response to a sinusoidal input $A \sin \omega t$, which can be written as a Fourier series:

$$\sum_{n=1}^{\infty} B_n \sin(n\omega t + \phi_n) \tag{19.35}$$

The **describing function** is the ratio of the complex Fourier coefficient $B_1 e^{j\phi_1}$ of the fundamental frequency of this output, to the amplitude A of the input. That is, the *describing function* is the complex function of ω, $(B_1/A)e^{j\phi_1}$, a frequency response function of an approximation of the nonlinear element. Thus the describing function represents the effective gain of the nonlinear element at the frequency of the input sinusoid.

In general, B_1 and ϕ_1 are functions of both the input frequency $\omega = 2\pi/T$ and the input amplitude A. Therefore we may write $B_1 = B_1(A, \omega)$, $\phi_1 = \phi_1(A, \omega)$ and the describing function as

$$\bar{N}(A, \omega) = \frac{B_1 e^{j\phi_1}}{A} = \frac{B_1(A, \omega)\, e^{j\phi_1(A,\omega)}}{A} \tag{19.36}$$

To apply the method, we replace system nonlinearities by describing functions and then apply the frequency domain techniques of Chapters 11, 12, and 15 through 18, with some modifications to account for the dependence of B_1 and ϕ_1 on A.

EXAMPLE 19.17. The output of the nonlinear function $f(e) = e^3$ in response to an input $e = A \sin \omega t$ is

$$f(e) = A^3 \sin^3 \omega t = \frac{A^3}{4}(3 \sin \omega t - \sin^3 \omega t)$$

From Equation (19.36), the describing function for $f(e)$ is

$$\bar{N}(A) = \frac{3A^2}{4}$$

Note that this nonlinearity produces no phase shift, so that $\phi_1(A, \omega) = 0$.

Hysteresis

A common type of nonlinearity called **hysteresis** or **backlash** is shown in Fig. 19-10. In electrical systems, it may occur due to nonlinear electromagnetic properties and, in mechanical systems, it may result from backlash in gear trains or mechanical linkages. For another example, see Problem 2.16.

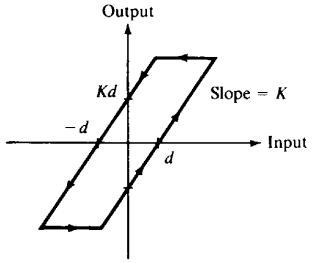


Fig. 19-10

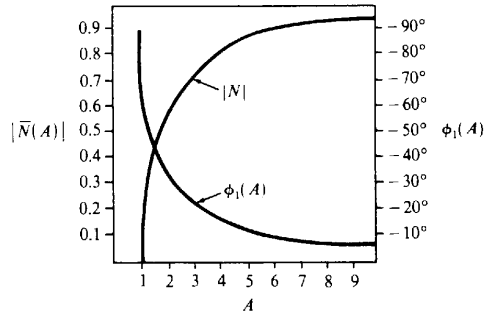


Fig. 19-11

The describing function characteristic for hysteresis, normalized to **dead zone** parameter $d = 1$ and slope $K = 1$, is shown in Fig. 19-11. The phase lag $\phi_1(A)$ of this describing function is a function of the input amplitude A, but is independent of the input frequency ω.

The describing function technique is particularly well suited for analysis of continuous or discrete-time systems containing a single nonlinear element, as illustrated in Fig. 19-12, with open-loop transfer function $GH = \bar{N}(A, \omega)G(\omega)$. Frequency response analysis of such systems typically entails first determining whether there exist values of A and ω that satisfy the characteristic equation, $1 + \bar{N}(A, \omega)G(\omega) = 0$, or

$$G(\omega) = -\frac{1}{\bar{N}(A, \omega)}$$

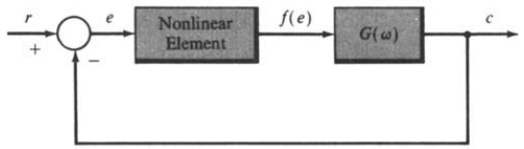


Fig. 19-12

that is, values of A and ω permitting oscillations. Nyquist, Bode, or Nichols chart plots of G and $-1/\bar{N}$ separately can be used to resolve this problem, because the plots must intersect if such A and ω exist. Relative stability can also be evaluated from such plots, by determining the additional gain (gain margin) and/or phase shift (phase margin) required to have the curves intersect.

It must be kept in mind that the describing function is only an approximation for the nonlinearity. The accuracy of describing function methods, using frequency response analysis based on linear system methods, depends upon the effective filtering by the plant $G(\omega)$ of the (neglected) higher than first-order harmonics produced by the nonlinearity. Since most plants have more poles than zeros, it is often a reasonable approximation.

EXAMPLE 19.18. Consider the system of Fig. 19-12 with $G(\omega) = 8/j\omega(j\omega + 2)^2$ and the saturation nonlinearity of Problem 19.17. Polar plots of $G(\omega)$ and $-1/\bar{N}(A)$ are shown in Fig. 19-13.

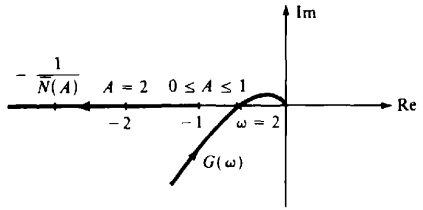


Fig. 19-13

There are no values of A and ω for which the two plots intersect, indicating that the system is stable and sustained oscillations of constant amplitude are not possible. However, if the forward-loop gain were increased by a factor of 2, from 8 to 16, the plots would intersect at $(-1,0)$ for $\omega = 2$ and $0 < A < 1$, and sustained oscillations would be possible. Thus an approximate gain margin for this system is 2 (6 db).

Popov's Stability Criterion

This criterion was developed for nonlinear feedback systems with a single nonlinear element in the loop, for example, as shown in Fig. 19-12. Such systems are stable if the linear element G is stable, $\mathrm{Re}\, G(\omega) > -1/K$, and the nonlinear element $f(e)$ satisfies the conditions: $f(0) = 0$ and $0 < f(e)/e < K$ for $e \neq 0$. Note that this criterion does not involve any approximations. Nyquist analysis is particularly well suited for its application.

EXAMPLE 19.19. For the system of Fig. 19-12, with $G = 1/(j\omega + 1)^3$, the Polar Plot is shown in Fig. 19-14. For all ω, $\mathrm{Re}\, G \ge -1/4$. Therefore the nonlinear system is stable if $K < 4$, $f(0) = 0$, and $0 < f(e)/e < K$ for $e \neq 0$.

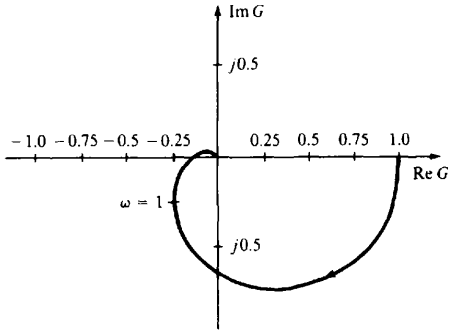


Fig. 19-14

EXAMPLE 19.20. For the nonlinear system in Fig. 19-12, with a stable discrete-time plant $G = 1/z$,

$$G(e^{j\omega T}) = e^{-j\omega T} = \cos \omega T - j \sin \omega T$$

The circular Polar Plot of G is shown in Fig. 19-15, and

$$\mathrm{Re}\, G(e^{j\omega T}) > \frac{-1}{K} \qquad \text{for} \quad K < 1$$

Thus the system is stable if $f(0) = 0$ and $0 < f(e)/e < K < 1$ for $e \neq 0$.

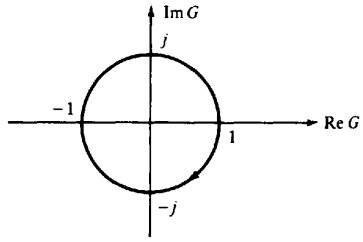


Fig. 19-15

Solved Problems

NONLINEAR CONTROL SYSTEMS

19.1. Several types of *control laws* or *control algorithms* were presented in Definitions 2.25 through 2.29. Which of these are nonlinear and which are linear, from the viewpoint of their input-output characteristics?

The on-off (binary) controller of Definition 2.25 is clearly nonlinear, its output being a discontinuous function of its input. The remaining controllers, that is, the proportional (P), derivative (D), integral (I) and PD, PI, DI, and PID controllers given in Definitions 2.26 through 2.29, are all linear. Each of their outputs are defined by linear operations, or linear combinations of linear operations, on each of their inputs.

19.2. Why is the thermostatically controlled heating system described in Problem 2.16 nonlinear?

The thermostat controller in this system is a nonlinear binary device, with a hysteresis input-output characteristic, as described in Problem 2.16. This controller regulates the room temperature output of this control system in an oscillatory manner between upper and lower limits bracketing the desired temperature setting. This type of behavior is characteristic of many nonlinear control systems.

LINEARIZED AND PIECEWISE-LINEAR SYSTEM APPROXIMATIONS

19.3. The differential equation of a certain physical system is given by

$$\frac{d^3y}{dt^3} + 4\frac{d^2y}{dt^2} + f(y) = 0$$

The function $f(y)$ is nonlinear, but it can be approximated by the piecewise-linear graph illustrated in Fig. 19-16. Determine a piecewise-linear approximation for the nonlinear system differential equation.

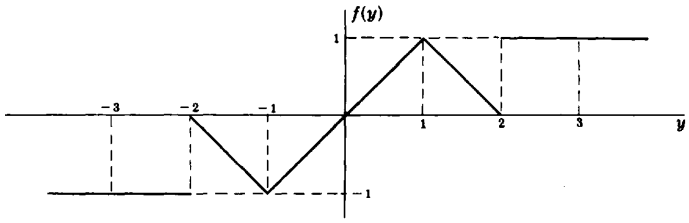


Fig. 19-16

The nonlinear system can be approximated by the following set of five linear equations over the indicated ranges of y:

$$\frac{d^3y}{dt^3} + 4\frac{d^2y}{dt^2} - 1 = 0 \qquad y < -2$$

$$\frac{d^3y}{dt^3} + 4\frac{d^2y}{dt^2} - y - 2 = 0 \qquad -2 \le y < -1$$

$$\frac{d^3y}{dt^3} + 4\frac{d^2y}{dt^2} + y = 0 \qquad -1 \le y \le 1$$

$$\frac{d^3y}{dt^3} + 4\frac{d^2y}{dt^2} - y + 2 = 0 \qquad 1 < y \le 2$$

$$\frac{d^3y}{dt^3} + 4\frac{d^2y}{dt^2} + 1 = 0 \qquad 2 < y$$

19.4. A solution of the nonlinear differential equation

$$\frac{d^2y}{dt^2} + y\cos y = u$$

with input $u = 0$, is $y = 0$. Linearize the differential equation about this input and output using a Taylor series expansion of the function $d^2y/dt^2 + y\cos y - u$ about the point $u = y = 0$.

The Taylor series expansion of $\cos y$ about $y = 0$ is

$$\cos y = \sum_{k=0}^{\infty} \frac{y^k}{k!}\left[\frac{d^k}{dy^k}(\cos y)\bigg|_{y=0}\right] = 1 - \frac{1}{2!}y^2 + \cdots$$

Therefore $$\frac{d^2y}{dt^2} + y\cos y - u = \frac{d^2y}{dt^2} + y\left(1 - \frac{y^2}{2!} + \cdots\right) - u$$

Keeping only first-degree terms, the linearized equation is $d^2y/dt^2 + y = u$. This equation is valid only for small deviations (perturbations) about the operating point $u = y = 0$.

19.5. Write the perturbation equations determined in Example 19.9 in vector-matrix form. Why are they linear? Under what conditions would they be time-invariant?

$$\frac{d(\Delta\mathbf{x})}{dt} \equiv \begin{bmatrix} \dfrac{d(\Delta x_1)}{dt} \\ \dfrac{d(\Delta x_2)}{dt} \end{bmatrix} \cong \begin{bmatrix} -2c_2\bar{x}_1(t) & c_1\bar{u}(t) \\ \dfrac{c_3c_4}{[c_4 + \bar{x}_1(t)]^2} & 0 \end{bmatrix} \Delta\mathbf{x} + \begin{bmatrix} c_1\bar{x}_2(t) \\ 0 \end{bmatrix} \Delta u$$

$$\Delta y \cong [2c_5\bar{x}_1(t) \quad 0]\,\Delta\mathbf{x}$$

These equations are linear because the matrices premultiplying $\Delta\mathbf{x}$ and Δu are independent of $\Delta\mathbf{x}$ and $\Delta\mathbf{u}$. They would be time-invariant if the parameters $c_1, c_2, \ldots, c_5$ were constant and the "operating point" of the system, for $u = \bar{u}(t)$ and $x = \bar{x}(t)$, were also constant. This would be the case if $\bar{u}$ = constant.

19.6. Derive the linearized Equations (*19.13*) and (*19.14*) for the nonlinear differential system given by (*19.11*) and (*19.12*).

We consider changes $\Delta\mathbf{x}$ in $\mathbf{x}$ as a result of changes $\Delta\mathbf{u}$ in $\mathbf{u}$, each about operating points $\bar{\mathbf{x}}$ and $\bar{\mathbf{u}}$, respectively, that is,

$$\mathbf{x}(t) = \bar{\mathbf{x}}(t) + \Delta\mathbf{x}(t)$$

$$\mathbf{u}(t) = \bar{\mathbf{u}}(t) + \Delta\mathbf{u}(t)$$

In these equations, t is considered a parameter, held constant in the derivation. We therefore suppress t, for convenience. Substitution of $\bar{\mathbf{x}} + \Delta\mathbf{x}$ for $\mathbf{x}$ and $\bar{\mathbf{u}} + \Delta\mathbf{u}$ for $\mathbf{u}$ in (*19.11*) gives

$$\frac{d\mathbf{x}}{dt} = \frac{d\bar{\mathbf{x}}}{dt} + \frac{d(\Delta\mathbf{x})}{dt} = \mathbf{f}(\bar{\mathbf{x}} + \Delta\mathbf{x}, \bar{\mathbf{u}} + \Delta\mathbf{u})$$

Now we expand this equation in a Taylor series about $\{\bar{\mathbf{x}}, \bar{\mathbf{u}}\}$, retaining only first-order terms:

$$\frac{d\mathbf{x}}{dt} + \frac{d(\Delta\mathbf{x})}{dt} \cong \mathbf{f}(\bar{\mathbf{x}}, \bar{\mathbf{u}}) + \frac{d\mathbf{f}}{\partial\mathbf{x}}\bigg|_{\substack{\mathbf{x}=\bar{\mathbf{x}}(t) \\ \mathbf{u}=\bar{\mathbf{u}}(t)}} \Delta\mathbf{x} + \frac{\partial\mathbf{f}}{\partial\mathbf{u}}\bigg|_{\substack{\mathbf{x}=\bar{\mathbf{x}}(t) \\ \mathbf{u}=\bar{\mathbf{u}}(t)}} \Delta\mathbf{u}$$

Then, since $d\bar{\mathbf{x}}/dt = f(\bar{\mathbf{x}}, \bar{\mathbf{u}})$, Equation (*19.11*) follows immediately after subtracting these corresponding terms from both sides of the equation above. Similarly, for

$$\mathbf{y} = \mathbf{g}(\mathbf{x})$$

$$\mathbf{y} = \bar{\mathbf{y}} + \Delta\mathbf{y} = \mathbf{g}(\bar{\mathbf{x}} + \Delta\mathbf{x}) \cong \mathbf{g}(\bar{\mathbf{x}}) + \frac{\partial\mathbf{g}}{\partial\mathbf{x}}\bigg|_{\mathbf{x}=\mathbf{x}} \Delta\mathbf{x} = \bar{\mathbf{y}} + \frac{\partial\mathbf{g}}{\partial\mathbf{x}}\bigg|_{\mathbf{x}=\bar{\mathbf{x}}} \Delta\mathbf{x}$$

Subtracting $\bar{\mathbf{y}}$ from both sides finally gives

$$\Delta\mathbf{y} \cong \frac{\partial\mathbf{g}}{\partial\mathbf{x}}\bigg|_{\mathbf{x}=\bar{\mathbf{x}}} \Delta\mathbf{x}$$

19.7. The equations describing the motion of an earth satellite in the orbit plane are

$$r\frac{d^2\theta}{dt^2} + 2\frac{dr}{dt}\frac{d\theta}{dt} = 0 \qquad \frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = -\frac{k^2}{pr^2}$$

(See Problem 3.3 and Example 19.2 for more details.) A satellite is in a nearly circular orbit determined by r and $d\theta/dt \equiv \omega$. An exactly circular orbit is defined by

$$r = r_0 = \text{constant} \qquad \omega = \omega_0 = \text{constant}$$

Since $dr_0/dt = 0$ and $d\omega_0/dt = 0$, the first differential equation is eliminated for a circular orbit. The second equation reduces to $r_0\omega_0^2 = k^2/pr_0^2$. Find a set of linear equations which approximately describes the differences

$$\delta r \equiv r - r_0 \qquad \delta\omega \equiv \omega - \omega_0$$

In the equations of motion we make the substitutions

$$r = r_0 + \delta r \qquad \omega = \omega_0 + \delta\omega$$

and obtain the equations

$$(r_0 + \delta r)\frac{d(\omega_0 + \delta\omega)}{dt} + 2\frac{d(r_0 + \delta r)}{dt}(\omega_0 + \delta\omega) = 0$$

$$\frac{d^2(r_0 + \delta r)}{dt^2} - (r_0 + \delta r)(\omega_0 + \delta\omega)^2 = -\frac{k}{p(r_0 + \delta r)^2}$$

We note that

$$\frac{d(r_0+\delta r)}{dt}=\frac{d(\delta r)}{dt} \qquad \frac{d^2(r_0+\delta r)}{dt^2}=\frac{d^2(\delta r)}{dt^2} \qquad \frac{d(\omega_0+\delta\omega)}{dt}=\frac{d(\delta\omega)}{dt}$$

since both r_0 and ω_0 are constant. The first differential equation then becomes

$$r_0\frac{d(\delta\omega)}{dt}+(\delta r)\frac{d(\delta\omega)}{dt}+2\omega_0\frac{d(\delta r)}{dt}+2\frac{d(\delta r)}{dt}\delta\omega=0$$

Since the differences δr, $\delta\omega$ and their derivatives are small, the second-order terms $(\delta r)(d(\delta\omega)/dt)$ and $2(d(\delta r)/dt)\,\delta\omega$ can be assumed negligible and eliminated. The resulting linear equation is

$$r_0\frac{d(\delta\omega)}{dt}+2\omega_0\frac{d(\delta r)}{dt}=0$$

which is one of the two desired equations. The second differential equation can be rewritten as

$$\frac{d^2(\delta r)}{dt^2}-r_0\omega_0^2-2r_0\omega_0\delta\omega-r_0(\delta\omega)^2-\omega_0^2\delta r-2\omega_0(\delta r)(\delta\omega)-(\delta\omega)^2\delta r$$
$$=-\frac{k}{pr_0^2}-\frac{2k\delta r}{r_0^3}+\text{higher-order terms in }\delta r\text{ and }\delta\omega$$

where the right-hand side is the Taylor series expansion of $-k/pr^2$ about r_0. All terms of order 2 and greater in δr and $\delta\omega$ may again be assumed negligible and eliminated leaving the linear equation

$$\frac{d^2(\delta r)}{dt^2}-r_0\omega_0^2-2r_0\omega_0\delta\omega-\omega_0^2\delta\omega-\omega_0^2\delta r=-\frac{k}{pr_0^2}-\frac{2k\delta r}{pr_0^3}$$

In the problem statement we saw that $r_0\omega_0^2=k/pr_0^2$. Hence the final equation is

$$\frac{d^2(\delta r)}{dt^2}-2r_0\omega_0\delta\omega-\omega_0^2\delta r=-\frac{2k\delta r}{pr_0^3}$$

which is the second of the two desired linearized equations.

PHASE PLANE METHODS

19.8. Show the equation $d^2x/dt^2=f(x,dx/dt)$ can be equivalently described by a pair of first-order differential equations.

We define a set of new variables: $x_1\equiv x$ and $x_2\equiv dx_1/dt=dx/dt$.

$$\frac{d^2x}{dt^2}=\frac{d^2x_1}{dt}=\frac{dx_2}{dt}=f\left(x,\frac{dx}{dt}\right)=f\left(x_1,\frac{dx_1}{dt}\right)=f(x_1,x_2)$$

The two desired equations are therefore

$$\frac{dx_1}{dt}=x_2 \qquad \frac{dx_2}{dt}=f(x_1,x_2)$$

19.9. Show that the phase plane trajectory of the solution of the differential equation

$$\frac{d^2x}{dt^2}+x=0$$

with initial conditions $x(0)=0$ and $(dx/dt)|_{t=0}=1$ is a circle of unit radius centered at the origin.

Letting $x\equiv x_1$ and $x_2\equiv dx_1/dt$, we obtain the pair of equations

$$\frac{dx_1}{dt}=x_2 \qquad x_1(0)=0$$
$$\frac{dx_2}{dt}=-x_1 \qquad x_2(0)=1$$

We eliminate time as the independent variable by writing

$$\frac{dx_1}{dx_2}=-\frac{x_2}{x_1} \quad \text{or} \quad x_1\,dx_1+x_2\,dx_2=0$$

Integrating this equation for the given initial conditions, we obtain

$$\int_0^{x_1}x_1'\,dx_1'+\int_1^{x_2}x_2'\,dx_2'=\tfrac{1}{2}x_1^2+\tfrac{1}{2}x_2^2-\tfrac{1}{2}=0 \quad \text{or} \quad x_1^2+x_2^2=1$$

which is the equation of a circle of unit radius centered at the origin.

19.10. Determine the equation of the phase plane trajectory of the equation

$$\frac{d^2x}{dt^2}+\frac{dx}{dt}=0$$

with the initial conditions $x(0)=0$ and $(dx/dt)|_{t=0}=1$.

With $x_1\equiv x$ and $x_2\equiv dx_1/dt$ we obtain the pair of first-order equations

$$\frac{dx_1}{dt}=x_2 \qquad x_1(0)=0$$
$$\frac{dx_2}{dt}=-x_2 \qquad x_2(0)=1$$

We eliminate time as the independent variable by writing

$$\frac{dx_1}{dx_2}=-\frac{x_2}{x_2}=-1 \quad \text{or} \quad dx_1+dx_2=0$$

Then

$$\int_0^{x_1}dx_1'+\int_1^{x_2}dx_2'=x_1+x_2-1=0 \quad \text{or} \quad x_1+x_2=1$$

which is the equation of a straight line, as shown in Fig. 19-17. The direction of the motion in the phase plane is indicated by the arrow and is determined by noting that, initially, $x_2(0)=1$; therefore $dx_1/dt>0$ and x_1 is increasing, and $dx_2/dt<0$ and x_2 is decreasing. The trajectory ends at the point $(x_1,x_2)=(1,0)$, where $dx_1/dt=dx_2/dt=0$ and thus motion terminates.

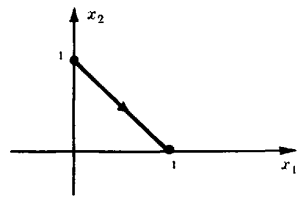


Fig. 19-17

19.11. Design an on-off controller for the system given by Equation (*19.19*) and Fig. 19-7, with $a = 0$.

For $a = 0$ in Equation (*19.19*), Equation (*19.23*) becomes

$$dx_1 = \frac{x_2 dx_2}{u}$$

The switching curve is generated by integrating this equation in the second quadrant with $u = +1$ and terminating at the origin, yielding

$$x_1(t) = -\frac{x_2^2(t)}{2} \qquad \text{or} \qquad e = -\frac{\dot{e}^2}{2}$$

and integrating in the fourth quadrant with $u = -1$ and terminating at the origin, yielding

$$x_1(t) = \frac{x_2^2(t)}{2} \qquad \text{or} \qquad e = \frac{\dot{e}^2}{2}$$

The switching curve is sketched in Fig. 19-18. The switching logic of this on-off controller is given in Table 19.2.

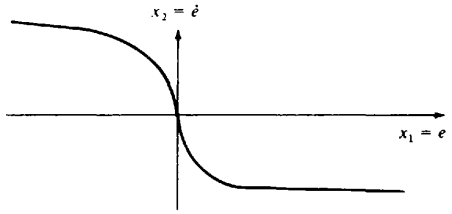


Fig. 19-18

Table 19.2

$\dot{e} > 0$	$e + \dot{e}^2/2 > 0$	$e - \dot{e}^2/2 > 0$	u
No	No	No	-1
No	No	Yes	$+1$
No	Yes	No	-1
No	Yes	Yes	$+1$
Yes	No	No	-1
Yes	No	Yes	-1
Yes	Yes	No	$+1$
Yes	Yes	Yes	$+1$

LYAPUNOV'S STABILITY CRITERION

19.12. Find the singular points of the pair of equations

$$\frac{dx_1}{dt} = \sin x_2 \qquad \frac{dx_2}{dt} = x_1 + x_2$$

Singular points are found by setting $\sin x_2 = 0$ and $x_1 + x_2 = 0$. The first equation is satisfied by $x_2 = \pm n\pi$, $n = 0, 1, 2, \ldots$. The second is satisfied by $x_1 = -x_2$. Hence the singular points are defined by

$$x_1 = \mp n\pi, \; x_2 = \pm n\pi \qquad n = 0, 1, 2, \ldots$$

19.13. The origin is a singular point for the pair of equations

$$\frac{dx_1}{dt} = ax_1 + bx_2 \qquad \frac{dx_2}{dt} = cx_1 + dx_2$$

Using Lyapunov theory, find sufficient conditions on a, b, c, and d such that the origin is asymptotically stable.

We choose a function

$$V = x_1^2 + x_2^2$$

which is positive for all x_1, x_2 except $x_1 = x_2 = 0$. The time derivative of V is

$$\frac{dV}{dt} = 2x_1\frac{dx_1}{dt} + 2x_2\frac{dx_2}{dt} = 2ax_1^2 + 2bx_1x_2 + 2cx_1x_2 + 2dx_2^2$$

To make dV/dt negative for all x_1, x_2, we might choose $a < 0$, $d < 0$, and $b = -c$. In this case,

$$\frac{dV}{dt} = 2ax_1^2 + 2dx_2^2 < 0$$

except when $x_1 = x_2 = 0$. Hence one set of sufficient conditions for asymptotic stability are $a < 0$, $d < 0$, and $b = -c$. There are other possible solutions to this problem.

19.14. Determine sufficient conditions for the stability of the origin of the nonlinear discrete-time system described by

$$x_1(k+1) = x_1(k) - f[x_1(k)]$$

Let $V[x(k)] = [x_1(k)]^2$, which is greater than 0 for all $x \neq 0$. Then

$$\Delta V = x_1^2(k+1) - x_1^2(k) = (x_1(k) - f[x_1(k)])^2 - x_1^2(k)$$
$$= -x_1(k)f[x_1(k)]\left(2 - \frac{f[x_1(k)]}{x_1}\right)$$

Therefore sufficient conditions for $\Delta V \leq 0$ and thus stability of the system are

$$x_1 f(x_1) \geq 0$$
$$\frac{f(x_1)}{x_1} \leq 2 \qquad \text{for all} \quad x_1$$

19.15. Determine sufficient conditions for the stability of the system

$$\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{b}f(x_1) \qquad \text{where} \quad A = \begin{bmatrix} -2 & -1 \\ 0 & -2 \end{bmatrix}, \; \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Let $V = \mathbf{x}^T P\mathbf{x}$ and $P = \begin{bmatrix} a & c \\ c & 1 \end{bmatrix}$. Then

$$\dot{V} = \mathbf{x}^T(PA + A^TP)\mathbf{x} + \mathbf{x}^TP\mathbf{b}f(x_1) + f(x_1)\mathbf{b}^TP\mathbf{x}$$
$$= \mathbf{x}^T\begin{bmatrix} -4a & -a-4c \\ -a-4c & -2c-4 \end{bmatrix}\mathbf{x} + 2(a+2c)x_1f(x_1) + 2(c+2)x_2f(x_1)$$

To eliminate the cross-product term $x_2 f(x_1)$, set $c = -2$. Then

$$\dot{V} = -\mathbf{x}^TQ\mathbf{x} + 2(a-4)x_1f(x_1)$$

where $Q \equiv \begin{bmatrix} 4a & a-8 \\ a-8 & 0 \end{bmatrix}$. For $Q \geq 0$, $a = 8$. The resulting $\dot{V}$ is

$$\dot{V} = -32x_1^2 + 8x_1f(x_1) = -8x_1^2\left(4 - \frac{f(x_1)}{x_1}\right)$$

Then $\dot{V} \leq 0$ and the system is stable if $f(x_1)/x_1 \leq 4$ for all $x_1 \neq 0$.

19.16. Determine sufficient conditions for stability of the nonlinear discrete-time system

$$\mathbf{x}(k+1) = A\mathbf{x}(k) + \mathbf{b}f[x_1(k)]$$

where $A = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$.

Let $V = \mathbf{x}^T P\mathbf{x}$, where $P = \begin{bmatrix} a & c \\ c & 1 \end{bmatrix}$. Then

$$\Delta V = V[\mathbf{x}(k+1)] - V[\mathbf{x}(k)] = \mathbf{x}(k+1)^T P\mathbf{x}(k+1) - \mathbf{x}(k)^T P\mathbf{x}(k)$$

$$= \left[f[x_1(k)]\mathbf{b}^T + \mathbf{x}(k)^T A^T\right] P\left[A\mathbf{x}(k) + \mathbf{b}f[x_1(k)]\right] - \mathbf{x}(k)^T P\mathbf{x}(k)$$

$$= \mathbf{x}^T(A^T PA - P)\mathbf{x} + f(x_1)\mathbf{b}^T P\mathbf{b}f(x_1) + f(x_1)\mathbf{b}^T PA\mathbf{x} + \mathbf{x}^T A^T P\mathbf{b}f(x_1)$$

where

$$A^T PA - P = \begin{bmatrix} 0 & a-2c \\ a-2c & a-2c \end{bmatrix} \quad \text{and} \quad \mathbf{b}^T PA = [-c \quad 1-c]$$

Now, in order for $A^T PA - P \le 0$, we set $a = 2c$ and, to eliminate the cross-product term $x_2 f(x_1)$, we set $c = 1$. Then $A^T PA - P = 0$ and

$$\Delta V = [f(x_1)]^2 - 2x_1 f(x_1) = -x_1 f(x_1)\left(2 - \frac{f(x_1)}{x_1}\right)$$

Sufficient conditions for $\Delta V \le 0$ and stability of the origin are then

$$x_1 f(x_1) \ge 0 \quad \text{and} \quad \frac{f(x_1)}{x_1} \le 2 \quad \text{for all} \quad x_1.$$

FREQUENCY RESPONSE METHODS

19.17. Show that the describing function for the piecewise-linear saturation element in Example 19.1 is given by

$$\frac{B_1}{A}e^{j\phi_1} = \frac{2}{\pi}\left[\sin^{-1}\frac{1}{A} + \frac{1}{A}\cos\sin^{-1}\frac{1}{A}\right]$$

We see from Fig. 19-1(b) that, when the magnitude of the input is less than 1.0, the output equals the input. When the input exceeds 1.0, then the output equals 1.0. Using the notation of Example 19.1, if

$$e(t) = A\sin\omega t \qquad A > 1$$

then $f(t)$ is as shown in Fig. 19-19 and can be written as

$$f(t) = \begin{cases} A\sin\omega t & \begin{cases} 0 \le t \le t_1 \\ t_2 \le t \le t_3 \\ t_4 \le t \le 2\pi/\omega \end{cases} \\ 1 & t_1 \le t \le t_2 \\ -1 & t_3 \le t \le t_4 \end{cases}$$

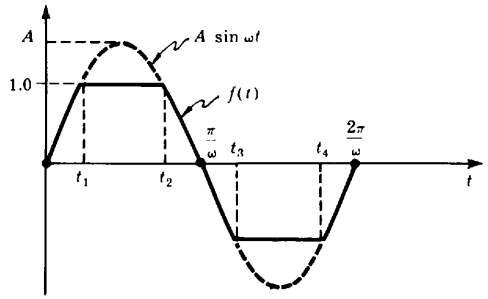


Fig. 19-19

The time t_1 is obtained by noting that

$$A\sin\omega t_1 = 1 \qquad \text{or} \qquad t_1 = \frac{1}{\omega}\sin^{-1}\frac{1}{A}$$

Similarly,

$$t_2 = \frac{\pi}{\omega} - \frac{1}{\omega}\sin^{-1}\frac{1}{A} \qquad t_3 = \frac{\pi}{\omega} + \frac{1}{\omega}\sin^{-1}\frac{1}{A} \qquad t_4 = \frac{2\pi}{\omega} - \frac{1}{\omega}\sin^{-1}\frac{1}{A}$$

The magnitude B_1 and phase angle ϕ_1 of the describing function are determined from the expression for the first Fourier coefficient:

$$B_1 = \frac{\omega}{\pi}\int_0^{2\pi/\omega} f(t)\sin\omega t\,dt$$

Since $f(t)$ is an odd function, the phase angle ϕ_1 is zero. The integral defining B_1 can be rewritten as

$$B_1 = \frac{\omega}{\pi}\int_0^{t_1} A\sin^2\omega t\,dt + \frac{\omega}{\pi}\int_{t_1}^{t_2}\sin\omega t\,dt$$
$$+ \frac{\omega}{\pi}\int_{t_2}^{t_3} A\sin^2\omega t\,dt - \frac{\omega}{\pi}\int_{t_3}^{t_4}\sin\omega t\,dt + \frac{\omega}{\pi}\int_{t_4}^{2\pi/\omega} A\sin^2\omega t\,dt$$

But

$$\int_0^{t_1} A\sin^2\omega t\,dt = \int_{t_4}^{2\pi/\omega} A\sin^2\omega t\,dt = \frac{1}{2}\int_{t_2}^{t_3} A\sin^2\omega t\,dt$$

and

$$\int_{t_1}^{t_2}\sin\omega t\,dt = -\int_{t_3}^{t_4}\sin\omega t\,dt = 2\int_{t_1}^{\pi/2\omega}\sin\omega t\,dt$$

We can thus write B_1 as

$$B_1 = \frac{4\omega}{\pi}\int_0^{t_1} A\sin^2\omega t\,dt + \frac{4\omega}{\pi}\int_{t_1}^{\pi/2\omega}\sin\omega t\,dt = \frac{2}{\pi}\left[A\omega t_1 - \frac{A}{2}\sin 2\omega t_1 + 2\cos\omega t_1\right]$$

Substituting $t_1 = (1/\omega)\sin^{-1}(1/A)$ and simplifying, we obtain

$$B_1 = \frac{2}{\pi}\left[A\sin^{-1}\frac{1}{A} + \cos\sin^{-1}\frac{1}{A}\right]$$

Finally, the describing function is

$$\frac{B_1}{A} = \frac{2}{\pi}\left[\sin^{-1}\frac{1}{A} + \frac{1}{A}\cos\sin^{-1}\frac{1}{A}\right]$$

19.18. Determine the amplitude A and frequency ω for which oscillations could be maintained in the system of Example 19.18 with the forward-loop gain increased to 32 from 8.

The Polar Plots of

$$G(\omega) = \frac{32}{j\omega(j\omega + 2)^2}$$

and $-1/\bar{N}(A)$ are shown in Fig. 19-20. The two loci intersect at $A = 2.5$ and $\omega = 2$, the conditions for oscillation.

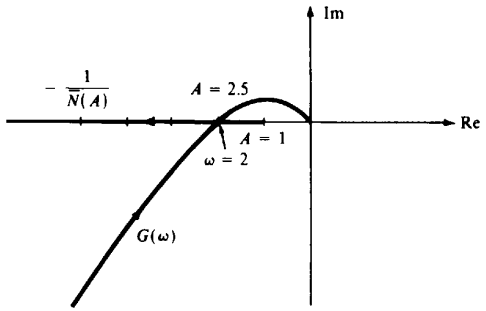


Fig. 19-20

19.19. Determine the amplitude and frequency of possible oscillations for the system of Fig. 19-12 with $f(e) = e^3$ and

$$G(\omega) = \frac{1}{(j\omega + 1)^3}$$

From Example 19.17, the describing function for this nonlinearity is

$$\bar{N}(A) = \frac{3A^2}{4} \quad \text{and} \quad -\frac{1}{\bar{N}} = -\frac{4}{3A^2}$$

From the Polar Plots shown in Fig. 19-21, $G(\omega)$ and $-1/\bar{N}$ intersect for $\omega = 1.732$ and $A = 3.27$, the conditions for oscillation.

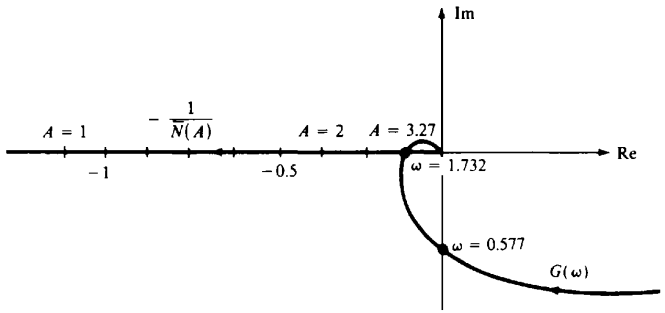


Fig. 19-21

19.20. Determine the amplitude and frequency of possible oscillations for the system of Fig. 19-12, with the hysteresis nonlinearity shown in Fig. 19-22, and $G(\omega) = 2/j\omega(j\omega + 1)$.

The system block diagram can be manipulated as shown in Fig. 19-23, so that the hysteresis element is normalized, with a dead zone of 1 and a slope of 1. Figure 19-11 can then be used to construct the Polar Plot of $-1/N$, shown in Fig. 19-24 with the Polar Plot of $2G(\omega)$, rather than $G(\omega)$, because the loop transfer function excluding the nonlinearity is $4G(\omega)/2 = 2G(\omega)$.

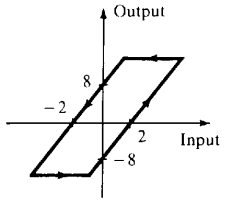


Fig. 19-22

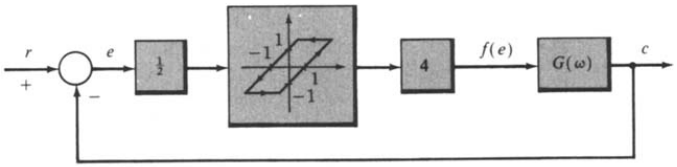


Fig. 19-23

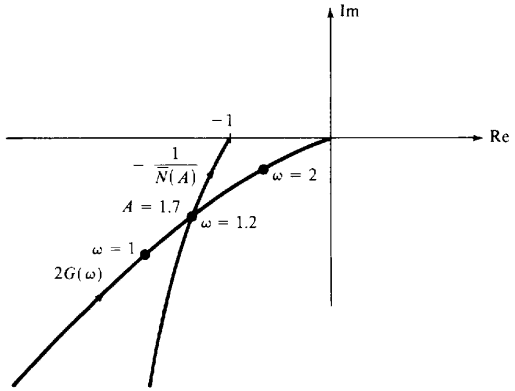


Fig. 19-24

The two curves intersect for $\omega = 1.2$ rad/sec and $A = 1.7$, the conditions for oscillation of the system. Note that A is the amplitude of the input to the normalized nonlinearity. Therefore the amplitude for oscillations is 3.4, in terms of e.

Supplementary Problems

19.21. Determine the phase plane trajectory of the solution of the differential equation

$$\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + 4x = 0$$

19.22. Using Lyapunov theory, find sufficient conditions on a_1 and a_0 which guarantee that the point $x = 0$, $dx/dt = 0$ is stable for the equation

$$\frac{d^2x}{dt^2} + a_1\frac{dx}{dt} + a_0x = 0$$

Chapter 20

Introduction to Advanced Topics in Control Systems Analysis and Design

20.1 INTRODUCTION

This final chapter is an introduction to advanced topics in control systems science. Each subject is discussed only briefly here to familiarize the reader with some of the terminology and mathematical level of advanced methodologies. It should also provide some of the motivation for advanced study. Time-domain state variable techniques, introduced in Chapters 3 and 4 and used extensively in Chapter 19, predominate in advanced methodological developments, mainly because they provide the basis for solving broader classes of control system problems, including far more complex problems than are amenable to frequency-domain methods.

20.2 CONTROLLABILITY AND OBSERVABILITY

Much of modern control theory is developed in the time domain, rather than the frequency domain, and the basic linear and time-invariant *plant* (controlled process) model is typically given a *state variable* description (Chapter 3), Equation (*3.25b*): $d\mathbf{x}(t)/dt = A\mathbf{x}(t) + B\mathbf{u}(t)$ for continuous system plants, or Equation (*3.36*): $\mathbf{x}(k+1) = A\mathbf{x}(k) + B\mathbf{u}(k)$ for discrete-time system plants. For either type of model, the output equation may be written as $\mathbf{y} = C\mathbf{x}$, where $\mathbf{y} = \mathbf{y}(t)$ or $\mathbf{y}(k)$, $\mathbf{x} = \mathbf{x}(t)$ or $\mathbf{x}(k)$, and C is a matrix of compatible dimension. We mention in passing that this basic model form is often used to represent *time-varying* linear systems, with matrices A, B, or C having time-varying elements, and (less often) *nonlinear* systems, with A, B, or C having elements that are functions of the state vector $\mathbf{x}$.

The concept of *controllability* addresses the question of whether it is possible to *control* or *steer* the state (vector) $\mathbf{x}$ from the input $\mathbf{u}$. Specifically, does there exist a physically realizable input $\mathbf{u}$ that can be applied to the plant over a finite period of time that will steer the entire state vector $\mathbf{x}$ (every one of the n components of $\mathbf{x}$) from any point $\mathbf{x}_0$ in *state space* to any other point $\mathbf{x}_1$? If yes, the plant is **controllable**; if no, it is **uncontrollable**.

The concept of *observability* is complementary to that of controllability. It addresses the question of whether it is possible to determine all of the n components of the state vector $\mathbf{x}$ by measurement of the output $\mathbf{y}$ over a finite period of time. If yes, the system is **observable**; if no, it is **unobservable**. Obviously, if $\mathbf{y} = \mathbf{x}$, that is, if all state variables are measured, the system is observable. However, if $\mathbf{y} \neq \mathbf{x}$ and C is not a square matrix, the plant may still be observable.

The controllability and observability properties of the plant have important practical consequences in analysis and, more importantly, design of modern feedback control systems. Intuitively, uncontrollable plants cannot be steered arbitrarily; and it is impossible to know all of the state variables of unobservable plants. These problems are clearly related, because together this means that unobservable states (or state variables) cannot be individually controlled if the control variable $\mathbf{u}$ is required to be a function of $\mathbf{x}$, that is, if *feedback* control is needed.

Linear, time-invariant plant models in state variable form [Equations (*3.25b*) or (*3.36*)] are **controllable** if and only if the following **controllability matrix** has rank n (n linearly independent columns), where n is the number of state variables in the state vector $\mathbf{x}$:

$$\begin{bmatrix} B & AB & A^2B & \cdots & A^{n-1}B \end{bmatrix} \tag{20.1}$$

Similarly, the plant model is **observable** if and only if the following **observability matrix** has

rank n (n linearly independent rows):

$$\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix} \tag{20.2}$$

EXAMPLE 20.1. Consider the following single-input single-output (SISO) plant model, with $\mathbf{x} \equiv \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and a_{11}, a_{12}, a_{22} each nonzero:

$$\frac{d\mathbf{x}}{dt} = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}u \qquad y = C\mathbf{x} = \begin{bmatrix} 1 & 0 \end{bmatrix}\mathbf{x}$$

To test if this model is controllable, we first evaluate the matrix given by Equation (*20.1*):

$$\begin{bmatrix} B & AB \end{bmatrix} = \begin{bmatrix} 1 & a_{11} \\ 0 & 0 \end{bmatrix}$$

By Definition 3.11, the two columns $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} a_{11} \\ 0 \end{bmatrix}$ would be linearly independent if the only constants α and β for which

$$\alpha\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \beta\begin{bmatrix} a_{11} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

where $\alpha \equiv \beta \equiv 0$. This is clearly *not* the case, because $\alpha = 1$ and $\beta = -1/a_{11}$ satisfies this equation. Therefore the two columns of $[B \quad AB]$ are linearly *dependent*, the rank of $[B \quad AB] = 1 \neq 2 = n$, and this plant is therefore *uncontrollable*.

Similarly, from Equation (*20.2*),

$$\begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ a_{11} & a_{12} \end{bmatrix}$$

For this matrix, the only α and β for which $\alpha[1 \quad 0] + \beta[a_{11} \quad a_{12}] = [0 \quad 0]$ are $\alpha \equiv \beta \equiv 0$, because $a_{12} \neq 0$. Therefore the rank of $\begin{bmatrix} C \\ CA \end{bmatrix}$ is $n = 2$ and this plant is *observable*.

20.3 TIME-DOMAIN DESIGN OF FEEDBACK SYSTEMS (STATE FEEDBACK)

Design of many feedback control systems may be accomplished using time-domain representations and the concepts of controllability and observability discussed above. As noted in earlier chapters, particularly Chapter 14, Root-Locus Design, linear control system design is often performed by manipulating the locations of the poles of the closed-loop transfer function (the roots of the characteristic equation), using appropriate compensators in the feedforward or feedback path to meet performance specifications. This approach is satisfactory in many circumstances, but it has certain limitations that can be overcome using a different design philosophy, called *state feedback* design, that permits *arbitrary* pole placement, thereby providing substantially more flexibility in design.

The basic idea underlying state feedback control system design is as follows for single-input continuous plants $d\mathbf{x}/dt = A\mathbf{x} + Bu$. The procedure is the same for discrete-time systems.

With reference to Fig. 2-1, we seek a state feedback control:

$$u = -G\mathbf{x} + r \tag{20.3}$$

where G is a $1 \times n$ feedback matrix of constant gains (to be designed) and r is the reference input. Combining these equations, the closed-loop system is given by

$$\frac{d\mathbf{x}}{dt} = (A - BG)\mathbf{x} + Br \tag{20.4}$$

If the plant is *controllable*, the matrix G exists that can yield any (arbitrary) set of desired roots for the characteristic equation of this closed-loop system, represented by $|\lambda I - A + BG| = 0$, where the λ solutions of this determinant equation are the roots. This is the basic result.

EXAMPLE 20.2. A block diagram of the state feedback system given by Equations (*20.3*) and (*20.4*) is shown in Fig. 20-1.

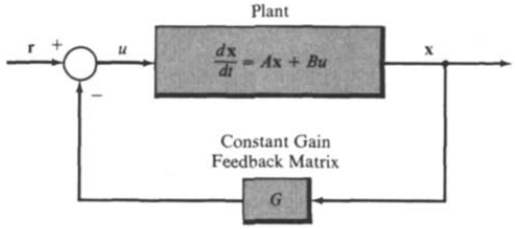


Fig. 20-1

To implement a state feedback design, the entire state vector $\mathbf{x}$ must somehow be made available, either as $\mathbf{x}$ exactly, or as an adequate approximation, denoted $\hat{\mathbf{x}}$. If the output is $\mathbf{y} = \mathbf{x}$, as in Fig. 20-1, there obviously is no problem. But, if all states are not available as outputs, which is more common, then *observability* of the plant model differential and output equations ($d\mathbf{x}/dt = A\mathbf{x} + B\mathbf{u}$ and $\mathbf{y} = C\mathbf{x}$) is required to obtain the needed state *estimate* or *observer* $\hat{\mathbf{x}}$. The equations for a typical state observer system are given by

$$\frac{d\hat{\mathbf{x}}}{dt} = (A - LC)\hat{\mathbf{x}} + L\mathbf{y} + B\mathbf{u} \qquad (20.5)$$

where A, B, and C are matrices of the plant and output measurement systems and L is an *observer design matrix* to be determined in a particular problem.

EXAMPLE 20.3. A detailed block diagram of the state observer system given by Equation (*20.5*) is shown in Fig. 20-2, along with the plant and measurement system block diagram (upper portion) for generating the needed input signals for the observer system (lower portion).

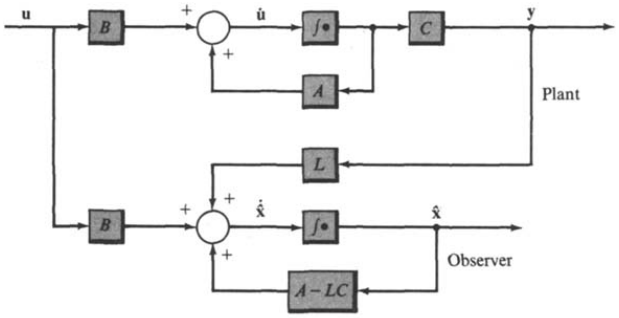


Fig. 20-2

EXAMPLE 20.4. Under suitable conditions, which include controllability and observability of the plant to be controlled, a *separation principle* applies and the state feedback portion (matrix G) and observer portion (matrix L) of a state feedback control system (with $\mathbf{y} \neq \mathbf{x}$) can be designed independently. A block diagram of the combined systems is shown in Fig. 20-3.

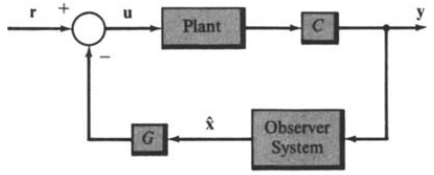


Fig. 20-3

We have omitted many details in this introductory material, and state feedback control systems are often more complex than described above.

20.4 CONTROL SYSTEMS WITH RANDOM INPUTS

System stimuli often include random or otherwise "unknown" components. This means that input functions may sometimes be more appropriately described probabilistically than deterministically. Such excitations are called **random processes.** System disturbances n (Definition 2.21), illustrated in several previous chapters, are sometimes represented by random process models in modern control theory and practice.

A random process can be viewed as a function of two variables, t and η, where t represents time and η a *random event*. The value of η is determined by chance.

EXAMPLE 20.5. A particular random process is denoted by $x(t, \eta)$. The random event η is the result of tossing an unbiased coin; heads or tails appears with equal probability. We define

$$x(t, \eta) \equiv \begin{cases} \text{a unit step function if } \eta = \text{heads} \\ \text{a unit ramp function if } \eta = \text{tails} \end{cases}$$

Thus $x(t, \eta)$ consists of two simple functions but is a random process because chance dictates which function occurs.

In practice, random processes consist of an infinity of possible time functions, called *realizations*, and we usually cannot describe them as explicitly as the one in Example 20.5. Instead, they must be described, in a statistical sense, by averages over all possible functions of time. The performance criteria discussed previously have all been related to specific inputs (e.g., K_P is defined for a unit step input, M_P and ϕ_{PM} for sine waves). But satisfaction of performance specifications defined for one input signal does not necessarily guarantee satisfaction for others. Therefore, for a random input, we cannot design for a *particular* signal, such as step function, but must design for the statistical average of random input signals.

EXAMPLE 20.6. The unit feedback system in Fig. 20-4 is excited by a random process input r having an infinity of possibilities. We want to determine compensation so that the error e is not excessive. There are an infinity of possibilities for r and, therefore, for e. Hence we cannot ask that each possible error satisfy given performance criteria but only that average errors be small. For instance, we might ask that G_1 be chosen from the set of all causal systems such that, as time goes to infinity, the statistical average of $e^2(t)$ does not exceed some constant, or is minimized.

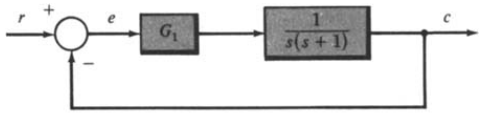


Fig. 20-4

The study of random processes in control systems, often called *stochastic control theory*, is an advanced level subject in applied mathematics.

20.5 OPTIMAL CONTROL SYSTEMS

The design problems discussed in earlier chapters are, in an elementary sense, optimal control problems. The classical measures of system performance such as steady state error, gain margin, and phase margin are essentially criteria of optimality, and control system compensators are designed to meet these requirements. In more general optimal control problems, the system measure of performance, or *performance index*, is not fixed beforehand. Instead, compensation is chosen so that the performance index is *maximized* or *minimized*. The value of the performance index is unknown until the completion of the optimization process.

In many problems, the performance index is a measure or function of the error $e(t)$ between the actual and ideal responses. It is formulated in terms of the design parameters chosen, to optimize the performance index, subject to existing physical constraints.

EXAMPLE 20.7. For the system illustrated in Fig. 20-5 we want to find a $K \geq 0$ such that the integral of the square of the error e is minimized when the input is a unit step function. Since $e \equiv e(t)$ is not constant, but a function of time, we can formulate this problem as follows: Choose $K \geq 0$ such that $\int_0^\infty e^2(t)\,dt$ is minimized, where

$$e(t) = \mathscr{L}^{-1}\left[\frac{s+2}{s^2+2s+K}\right] = \sqrt{\frac{K}{K-1}}\, e^{-t}\sin\left(\sqrt{K-1}\,t + \tan^{-1}\sqrt{K-1}\right)$$

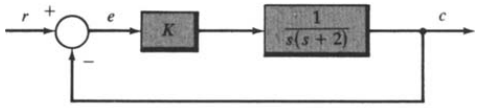


Fig. 20-5

The solution may be obtained for $K > 1$ using conventional minimization techniques of integral calculus, as follows:

$$\int_0^\infty e^2(t)\,dt = \frac{K}{K-1}\int_0^\infty \left[e^{-t}\sin\left(\sqrt{K-1}\,t + \tan^{-1}\sqrt{K-1}\right)\right]^2 dt$$

Integration yields

$$\int_0^\infty e^2(t)\,dt = \left(\frac{K}{K-1}\right)\left(\frac{e^{-2t}}{4}\right)\left[-1 - \frac{\cos\left(2\sqrt{K-1}\,t + 2\tan^{-1}\sqrt{K-1} - \tan^{-1}\left(-\sqrt{K-1}\right)\right)}{\sqrt{K}}\right]\Bigg|_0^\infty$$

$$= \frac{K}{4(K-1)}\left[1 + \frac{\cos\left(2\tan^{-1}\sqrt{K-1} - \tan^{-1}\left(-\sqrt{K-1}\right)\right)}{K}\right]$$

But

$$\cos\left(2\tan^{-1}\sqrt{K-1} - \tan^{-1}\left(-\sqrt{K-1}\right)\right) = -\cos 3\sqrt{K-1} = 3\cos\sqrt{K-1} - 4\cos^3\sqrt{K-1}$$

$$= \frac{3K-4}{K\sqrt{K}}$$

Therefore

$$\int_0^\infty e^2(t)\,dt = \frac{K}{4(K-1)}\left(1 + \frac{3K-4}{K^2}\right) = \frac{K}{4(K-1)}\,\frac{(K-1)(K+4)}{K^2} = \frac{K+4}{4K}$$

The first derivative of $\int_0^\infty e^2(t)\,dt$ with respect to K is given by

$$\frac{d}{dK}\left(\frac{K+4}{4K}\right) = -\frac{1}{K^2}$$

Apparently, $\int_0^\infty e^2(t)\,dt$ decreases monotonically as K increases. Therefore the optimal value of K is $K = \infty$, which is of course unrealizable. For this value of K,

$$\lim_{K\to\infty}\int_0^\infty e^2(t)\,dt = \lim_{K\to\infty}\left(\frac{K+4}{4K}\right) = \frac{1}{4}$$

Note also that the natural frequency ω_n of the optimal system is $\omega_n = \sqrt{K} = \infty$ and the damping ratio $\xi = 1/\omega_n = 0$, making it marginally stable. Therefore only a *suboptimal* (less than optimal) system can be practically realized and its design depends on the specific application.

Typical optimal control problems, however, are much more complex than this simple example and they require more sophisticated mathematical techniques for their solution. We do little more here than mention their existence.

20.6 ADAPTIVE CONTROL SYSTEMS

In some control systems, certain parameters are either not constant, or they vary in an unknown manner. In Chapter 9 we illustrated one way of minimizing the effects of such contingencies by designing for minimum sensitivity. If, however, parameter variations are large or very rapid, it may be desirable to design for the capability of continuously measuring them and changing the compensation so that system performance criteria are always satisfied. This is called *adaptive control* design.

EXAMPLE 20.8. Figure 20-6 depicts an example block diagram of an adaptive control system. The parameters A and B of the plant are known to vary with time. The block labeled "Identification and Parameter Adjustment" continuously measures the input $u(t)$ and output $c(t)$ of the plant to *identify* (quantify) the parameters A and B. In this manner, a and b of the lead compensator are modified by the output of this element to satisfy system specifications. The design of the Identification and Parameter Adjustment block is the major problem of adaptive control, another subject requiring advanced knowledge of applied mathematics.

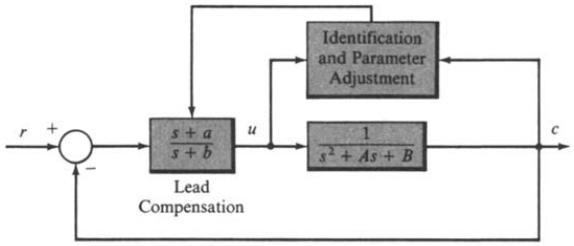


Fig. 20-6

Appendix A

Some Laplace Transform Pairs Useful for Control Systems Analysis

$F(s)$	$f(t)$ $t > 0$
1	$\delta(t)$ unit impulse
e^{-Ts}	$\delta(t-T)$ delayed impulse
$\frac{1}{s+a}$	e^{-at}
$\frac{1}{(s+a)^n}$	$\frac{1}{(n-1)!}t^{n-1}e^{-at}$ $n = 1,2,3,\ldots$
$\frac{1}{(s+a)(s+b)}$	$\frac{1}{b-a}(e^{-at}-e^{-bt})$
$\frac{s}{(s+a)(s+b)}$	$\frac{1}{a-b}(ae^{-at}-be^{-bt})$
$\frac{s+z_1}{(s+a)(s+b)}$	$\frac{1}{b-a}[(z_1-a)e^{-at}-(z_1-b)e^{-bt}]$
$\frac{1}{(s+a)(s+b)(s+c)}$	$\frac{e^{-at}}{(b-a)(c-a)}+\frac{e^{-bt}}{(c-b)(a-b)}+\frac{e^{-ct}}{(a-c)(b-c)}$
$\frac{s+z_1}{(s+a)(s+b)(s+c)}$	$\frac{(z_1-a)e^{-at}}{(b-a)(c-a)}+\frac{(z_1-b)e^{-bt}}{(c-b)(a-b)}+\frac{(z_1-c)e^{-ct}}{(a-c)(b-c)}$
$\frac{\omega}{s^2+\omega^2}$	$\sin\omega t$
$\frac{s}{s^2+\omega^2}$	$\cos\omega t$
$\frac{s+z_1}{s^2+\omega^2}$	$\sqrt{\frac{z_1^2+\omega^2}{\omega^2}}\sin(\omega t+\phi)$ $\phi \equiv \tan^{-1}(\omega/z_1)$
$\frac{s\sin\phi+\omega\cos\phi}{s^2+\omega^2}$	$\sin(\omega t+\phi)$
$\frac{1}{(s+a)^2+\omega^2}$	$\frac{1}{\omega}e^{-at}\sin\omega t$

$F(s)$	$f(t)$ $t > 0$
$\frac{1}{s^2+2\zeta\omega_n s+\omega_n^2}$	$\frac{1}{\omega_d}e^{-\zeta\omega_n t}\sin\omega_d t$ $\omega_d \equiv \omega_n\sqrt{1-\zeta^2}$
$\frac{s+a}{(s+a)^2+\omega^2}$	$e^{-at}\cos\omega t$
$\frac{s+z_1}{(s+a)^2+\omega^2}$	$\sqrt{\frac{(z_1-a)^2+\omega^2}{\omega^2}}e^{-at}\sin(\omega t+\phi)$ $\phi \equiv \tan^{-1}\left(\frac{\omega}{z_1-a}\right)$
$\frac{1}{s}$	$\mathbf{1}(t)$ unit step
$\frac{1}{s}e^{-Ts}$	$\mathbf{1}(t-T)$ delayed step
$\frac{1}{s}(1-e^{-Ts})$	$\mathbf{1}(t)-\mathbf{1}(t-T)$ rectangular pulse
$\frac{1}{s(s+a)}$	$\frac{1}{a}(1-e^{-at})$
$\frac{1}{s(s+a)(s+b)}$	$\frac{1}{ab}\left(1-\frac{be^{-at}}{b-a}+\frac{ae^{-bt}}{b-a}\right)$
$\frac{s+z_1}{s(s+a)(s+b)}$	$\frac{1}{ab}\left(z_1-\frac{b(z_1-a)e^{-at}}{b-a}+\frac{a(z_1-b)e^{-bt}}{b-a}\right)$
$\frac{1}{s(s^2+\omega^2)}$	$\frac{1}{\omega^2}(1-\cos\omega t)$
$\frac{s+z_1}{s(s^2+\omega^2)}$	$\frac{z_1}{\omega^2}-\sqrt{\frac{z_1^2+\omega^2}{\omega^4}}\cos(\omega t+\phi)$ $\phi \equiv \tan^{-1}(\omega/z_1)$
$\frac{1}{s(s^2+2\zeta\omega_n s+\omega_n^2)}$	$\frac{1}{\omega_n^2}-\frac{1}{\omega_n\omega_d}e^{-\zeta\omega_n t}\sin(\omega_d t+\phi)$ $\omega_d \equiv \omega_n\sqrt{1-\zeta^2}$ $\phi \equiv \cos^{-1}\zeta$
$\frac{1}{s(s+a)^2}$	$\frac{1}{a^2}(1-e^{-at}-ate^{-at})$
$\frac{s+z_1}{s(s+a)^2}$	$\frac{1}{a^2}[z_1-z_1e^{-at}+a(a-z_1)te^{-at}]$
$\frac{1}{s^2}$	t unit ramp
$\frac{1}{s^2(s+a)}$	$\frac{1}{a^2}(at-1+e^{-at})$
$\frac{1}{s^n}$ $n = 1,2,3,\ldots$	$\frac{t^{n-1}}{(n-1)!}$ $0! = 1$

Appendix B

Some *z*-Transform Pairs Useful for Control Systems Analysis

$F(z)$	k th term of time sequence $f(k)$, $k = 0, 1, 2, \ldots$
z^{-k}	1 at k, 0 elsewhere (Kronecker delta sequence)
$\frac{z}{z - e^{-aT}}$	e^{-akT}
$\frac{Te^{-aT}z}{(z - e^{-aT})^2}$	kTe^{-akT}
$\frac{T^2e^{-aT}z(z + e^{-aT})}{(z - e^{-aT})^3}$	$(kT)^2e^{-akT}$
$\frac{z^n}{(z - A)^n}$	$\frac{(k+1)(k+2)\cdots(k+n-1)}{(n-1)!}A^k$ (A is any complex number)
$\frac{z}{z - 1}$	1 (unit step sequence)
$\frac{Tz}{(z-1)^2}$	kT (unit ramp sequence)
$\frac{T^2z(z+1)}{(z-1)^3}$	$(kT)^2$
$\frac{z^n}{(z-1)^n}$	$\frac{(k+1)(k+2)\cdots(k+n-1)}{(n-1)!}$
$\frac{z\sin\omega T}{z^2 - 2z\cos\omega T + 1}$	$\sin\omega kT$
$\frac{z(z - \cos\omega T)}{z^2 - 2z\cos\omega T + 1}$	$\cos\omega kT$
$\frac{ze^{-aT}\sin\omega T}{z^2 - 2ze^{-aT}\cos\omega T + e^{-2aT}}$	$e^{-akT}\sin\omega kT$
$\frac{z(z - e^{-aT}\cos\omega T)}{z^2 - 2ze^{-aT}\cos\omega T + e^{-2aT}}$	$e^{-akT}\cos\omega kT$
$\frac{1}{(z-a)(z-b)}$	0 for $k = 0$ $\frac{1}{a-b}(a^{k-1} - b^{k-1})$ for $k > 0$
$\frac{z}{(z-a)(z-b)}$	$\frac{1}{a-b}(a^k - b^k)$
$\frac{z(1-a)}{(z-1)(z-a)}$	$1 - a^k$

References and Bibliography

1. Churchill, R. V. and Brown; J. W., *Complex Variables and Applications*, Fourth Edition, McGraw-Hill, New York, 1984.
2. Hartline, H. K. and Ratliff, F., "Inhibitory Interaction of Receptor Units in the Eye of the Limulus," *J. Gen. Physiol.*, 40:357, 1957.
3. Bliss, J. C. and Macurdy, W. B., "Linear Models for Contrast Phenomena," *J. Optical Soc. America*, 51:1375, 1961.
4. Reichardt, W. and MacGinitie, "On the Theory of Lateral Inhibition," *Kybernetic* (German), 1:155, 1962.
5. Desoer, C. A., "A General Formulation of the Nyquist Criterion," *IEEE Transactions on Circuit Theory*, Vol. CT-12, No. 2, June 1965.
6. Krall, A. M., "An Extension and Proof of the Root-Locus Method," *Journal of the Society for Industrial and Applied Mathematics*, Vol. 9, No. 4, December 1961, pp. 644–653.
7. Wiberg, D. M., *State Space and Linear Systems*, Schaum Outline Series, McGraw-Hill, New York, 1971.
8. LaSalle, J. and Lefschetz, S., *Stability by Liapunov's Direct Method, with Applications*, Academic Press, New York, 1958.
9. Lindorff, D. P., *Theory of Sampled-Data Control Systems*, John Wiley & Sons, New York, 1965.
10. Åström, K. J. and Wittenmark, B., *Computer Controlled Systems*, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.
11. Leigh, J. R., *Applied Digital Control*, Prentice-Hall, Englewood Cliffs, New Jersey, 1985.
12. Chen, C. T., *Introduction to Linear System Theory*, Second Edition, Holt, Rinehart and Winston, New York, 1985.
13. Truxal, J. G., *Automatic Feedback Control System Synthesis*, McGraw-Hill, New York, 1955.
14. Aizerman, M. A., *Theory of Automatic Control*, Addison-Wesley, Reading, Massachusetts, 1963.
15. Bode, H. W., *Network Analysis and Feedback Amplifier Design*, Van Nostrand, Princeton, New Jersey, 1945.
16. Brown, G. S. and Campbell, D. P., *Principles of Servomechanisms*, John Wiley, New York, 1948.
17. James, H. M., Nichols, N. B. and Phillips, R. S., *Theory of Servomechanisms*, McGraw-Hill, New York, 1947.
18. Kuo, B. C., *Automatic Control Systems*, Fifth Edition, Prentice-Hall, Englewood Cliffs, New Jersey 1987.

This page intentionally left blank

Appendix C

SAMPLE Screens From The Companion *Interactive Outline*

As described on the back cover, this book has a companion *Interactive Schaum's Outline* using Mathcad® which is designed to help you learn the subject matter more quickly and effectively. The *Interactive Outline* uses the LIVE-MATH environment of Mathcad technical calculation software to give you on-screen access to approximately 100 representative solved problems from this book, along with summaries of key theoretical points and electronic cross-referencing and hyperlinking. The following pages reproduce a representative sample of screens from the *Interactive Outline* and will help you understand the powerful capabilities of this electronic learning tool. Compare these screens with the associated solved problems from this book (the corresponding page numbers are listed at the start of each problem) to see how one complements the other.

In the *Interactive Schaum's Outline*, you'll find all related text, diagrams, and equations for a particular solved problem together on your computer screen. As you can see on the following pages, all the math appears in familiar notation, including units. The format differences you may notice between the printed *Schaum's Outline* and the *Interactive Outline* are designed to encourage your interaction with the material or show you alternate ways to solve challenging problems.

As you view the following pages, keep in mind that every number, formula, and graph shown ***is completely interactive when viewed on the computer screen.*** You can change the starting parameters of a problem and watch as new output graphs are calculated before your eyes; you can change any equation and immediately see the effect of the numerical calculations on the solution. Every equation, graph, and number you see is available for experimentation. Each adapted solved problem becomes a worksheet you can modify to solve dozens of related problems. The companion *Interactive Outline* thus will help you to learn and retain the material taught in this book more effectively and can also serve as a working problem-solving tool.

The Mathcad icon shown on the right is printed throughout this *Schaum's Outline*, indicating which problems are included in the *Interactive Outline*.

For more information about system requirements and the availability of titles in *Schaum's Interactive Outline Series*, please see the back cover.

Stability of Discrete-Time Systems

(Schaum's Feedback and Control Systems, 2nd ed., Solved Problem 5.22, pp. 124 - 125)

Statement

Is the system with the following characteristic equation stable?

$$z^4 + 2\cdot z^3 + 3\cdot z^2 + z + 1.0 = 0$$

System Parameters

In this problem, a numerical root-finding method is used which is justified in detail in **Appendix D**. For now, you may want to just follow along, and concentrate on the stability question. Create a vector of the polynomial coefficients, up to the **n – 1** power in the equation, starting with the zeroth-order term.

$$\text{coeff} := \begin{bmatrix} 1.0 \\ 1 \\ 3 \\ 2 \end{bmatrix} \quad \begin{matrix} \mathbf{1.0} \\ \mathbf{z} \\ \mathbf{z^2} \\ \mathbf{z^3} \end{matrix}$$

Coefficient of the **n**th power term: $A := 1$

Solution

Find the number of coefficients in the vector, and create a subdiagonal matrix of ones.

$$n := 0 .. \text{length}(\text{coeff}) - 2 \qquad C_{n+1,n} := 1$$

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

This is the subdiagonal matrix. For more information on the range variables and matrix functions used here, see **A Mathcad Tutorial**.

Solve for the eigenvalues of the matrix constructed from **C** and the coefficient vector. These are the roots of the equation.

$$Z := \text{eigenvals}\left(\text{augment}\left(C, -\frac{\text{coeff}}{A}\right)\right)$$

The roots of the equation are

$$Z = \begin{bmatrix} -0.043 + 0.641i \\ -0.043 - 0.641i \\ -0.957 + 1.227i \\ -0.957 - 1.227i \end{bmatrix}$$

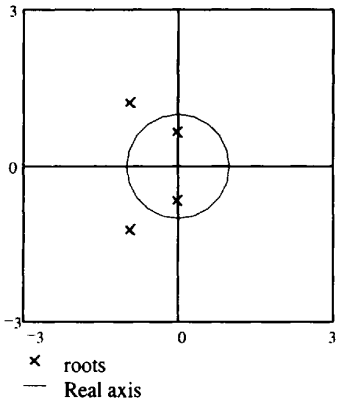
In order to graph these solutions, index them with a range variable:

$$i := 0 .. \text{length}(Z) - 1$$

Since this is a discrete-time system, the stability requirement is that the roots lie inside the unit circle. This will be graphed parametrically using sines and cosines, so define the range for θ.

$$\theta := 0, 0.1\cdot\pi .. 2\cdot\pi$$

The **z**-plane diagram for this system is



Because not all the roots are inside the unit circle, the system is unstable. You should take time to carefully examine the numerical root-finding technique shown here; it will be used throughout this Electronic Book.

Also, try changing the numbers in the vector of coefficients, **coeff**. See what sorts of discrete-time systems are stable. Can you find one? Can you find one that's marginally stable? What happens when you change the coefficient of the **nth** term, **A**?

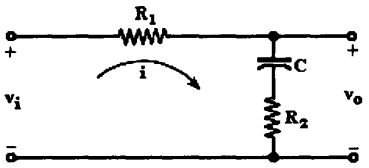
Lag Compensator

(Schaum's Feedback and Control Systems, 2nd ed., Solved Problems 6.13 and 6.16, pp. 138 and 139)

Statement

(a) Derive the transfer function of the R-C network implementation of the lag compensator shown in the figure below. (b) Derive the transfer function of two simple lag networks connected in cascade.

System Parameters



$\Omega \equiv \text{ohm}$

$\mu F \equiv 10^{-6} \cdot \text{farad}$

In order to graph the results, we use the following specific circuit element values, which are defined globally with the graphs below.

$R_1 = 200 \cdot \Omega \qquad R_2 = 50 \cdot \Omega \qquad C = 25 \cdot \mu F$

Solution

Kirchhoff's voltage law and constitutive relationships (Ohm's law) for the loop yield the equation

$$i \cdot R_1 + \frac{1}{C} \cdot \int_0^t i\,dt + i \cdot R_2 \equiv v_i$$

(a)

assuming zero initial conditions. Taking the Laplace transform of these two equations results in the equation

$$\left(R_1 + R_2 + \frac{1}{C \cdot s}\right) \cdot I(s) \equiv V_i(s)$$

Notice that this is the same expression which would have resulted if we used the expression **1/(C•s)** for the impedance of the capacitor in Kirchhoff's voltage law. Since the transfer function is the ratio of the output to the input, find $\mathbf{P_{LAG}(s) = V_o(s)/V_i(s)}$:

$$V_o \equiv \left(R_2 + \frac{1}{C \cdot s}\right) \cdot I(s)$$

This gives

$$P_{LAG}(s) \equiv \frac{R_2 + \frac{1}{C \cdot s}}{R_1 + R_2 - \frac{1}{C \cdot s}}$$

Compare this to the definition of a lag compensator given in **Chapter 6**. Here,

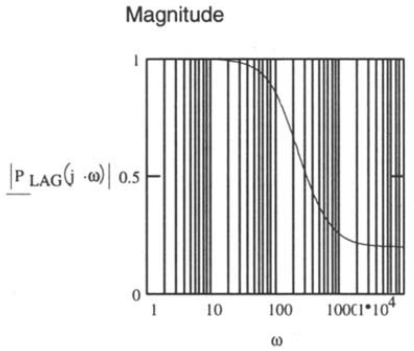
$$a \equiv \frac{1}{(R_1 + R_2) \cdot C} \qquad b \equiv \frac{1}{R_2 \cdot C}$$

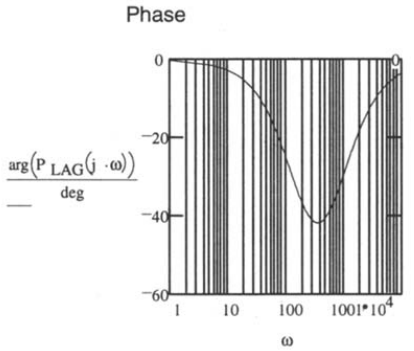
where **-a** is the pole of the system. To graph the frequency response, define a suitable range for ω.

$$\omega \equiv 1 \cdot \frac{rad}{sec}, 5 \cdot \frac{rad}{sec} .. 10^4 \cdot \frac{rad}{sec}$$

By changing the values of circuit elements below, examine their effect on the characteristics of the frequency response curves shown in the figures.

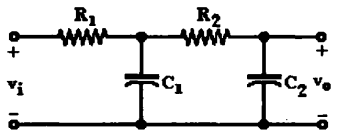
$R_1 \equiv 200 \cdot \Omega \qquad R_2 \equiv 50 \cdot \Omega \qquad C \equiv 25 \cdot \mu F$





Notice that these graphs are in semilog scale. We can see that this is a lag compensator from the graph of the phase: the response lags the input for all frequencies. It is possible to examine the simple lag network by setting $\mathbf{R_2}$ equal to zero.

(b) Suppose we examine the situation in which two simple lag networks are connected in cascade.



Using a voltage divider and the Laplace transform expression for the impedances,

$$V_2 = \frac{V_i \cdot \left[\frac{1}{R_2 + \frac{1}{C_2 \cdot s}} + \frac{1}{\left(\frac{1}{C_1 \cdot s}\right)}\right]^{-1}}{R_1 + \left[\frac{1}{R_2 + \frac{1}{C_2 \cdot s}} + \frac{1}{\left(\frac{1}{C_1 \cdot s}\right)}\right]^{-1}}$$

simplifies to

$$V_2 = \frac{V_i \cdot (R_2 \cdot C_2 \cdot s + 1)}{s^2 \cdot R_1 \cdot C_1 \cdot R_2 \cdot C_2 + (R_1 \cdot C_1 + R_2 \cdot C_2 + R_1 \cdot C_2) \cdot s + 1}$$

Using a second voltage divider, we obtain

$$V_0 = \frac{V_2 \cdot \frac{1}{C_2 \cdot s}}{R_2 + \frac{1}{C_2 \cdot s}} = \frac{V_2}{(R_2 \cdot C_2 \cdot s + 1)}$$

which, after expansion, becomes

$$V_0 = \frac{V_i}{s^2 \cdot R_1 \cdot C_1 \cdot R_2 \cdot C_2 + (R_1 \cdot C_1 + R_2 \cdot C_2 + R_1 \cdot C_2) \cdot s + 1}$$

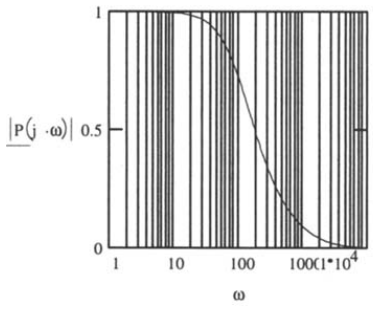
The expression for V_0 results in the transfer function

$$P(s) = \frac{1}{s^2 \cdot R_1 \cdot C_1 \cdot R_2 \cdot C_2 + (R_1 \cdot C_1 + R_2 \cdot C_2 + R_1 \cdot C_2) \cdot s + 1}$$

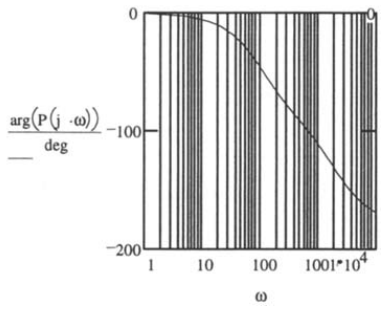
Experiment with the two capacitor values to see their effect on the lag compensator output.

$C_1 := 25 \cdot \mu F$ $C_2 := 20 \cdot \mu F$

Magnitude



Phase



Compare the single-stage and two-stage simple lag compensators:

$$P_{single}(s) := \frac{1}{(R_1 \cdot C \cdot s + 1)}$$

$$P_{double}(s) := \frac{1}{s^2 \cdot R_1 \cdot C_1 \cdot R_2 \cdot C_2 + (R_1 \cdot C_1 + R_2 \cdot C_2 + R_1 \cdot C_2) \cdot s + 1}$$

The second-order pole on the two-stage compensator greatly increases the amount of lag achieved in phase. Think about how you would use this information to best implement a compensator. Is the two-stage system stable? What would you do if you wished to add lag to a circuit operating at higher frequencies (notice that the response is almost zero at 1000 rad/sec)?

Frequency vs. Time-Domain Specifications

(Schaum's Feedback and Control Systems, 2nd ed., Example Problems 10.2 and 10.3, pp. 233 - 235)

Statement

Using the second order system shown first in **Chapter 3**, compare the frequency and time-domain specifications and plots.

System Parameters

$\omega_n := 100 \cdot \frac{rad}{sec}$ $\zeta := 0.2$ $dB := 1$

(These parameters are defined globally next to the graphs at the end of the problem, so you may experiment with them and watch the change in the graphs simultaneously.)

Solution

Beginning with the frequency-domain, examine the resonant peak, the cutoff frequency, and the bandwidth. The equation for the magnitude of the impulse response of the canonical second-order system is

$$Y(s) := \frac{\omega_n^2}{s^2 + 2 \cdot \zeta \cdot \omega_n \cdot s + \omega_n^2}$$

The magnitude of the response, in dB, is

$$MAG(\omega) := 20 \cdot \log(|Y(j \cdot \omega)|)$$

To find the peak value, take the derivative, as was done in **Chapter 10**.

$D(\omega) := \frac{d}{d\omega} |Y(j \cdot \omega)|$ Guess: $\omega := if\left(|\zeta| > .5, \frac{\omega_n}{2}, \omega_n\right)$

Find the frequency at which the derivative is zero.

$\omega_p := |root(D(\omega), \omega)|$ $\omega_p = 95.915 \cdot \frac{rad}{sec}$

Check: $D(\omega_p) = 8.938 \cdot 10^{-6} \cdot sec$

This is very close to zero, so ω_p is a good approximation of the resonant frequency.

The magnitude of the resonance peak is given by

$M_p := |Y(j \cdot \omega_p)|$ $M_p = 2.552$

The magnitude of the peak could be used to calculate the bandwidth, but since this is a lowpass system, it's probably best to base the bandwidth calculation on the value of the transfer function at dc.

In this case,

$\left|Y\left(0 \cdot \frac{rad}{sec}\right)\right| = 1$ or, in decibels, $MAG\left(0 \cdot \frac{rad}{sec}\right) = 0 \cdot dB$

$$\omega_c := \left| root\left[|Y(j \cdot \omega)| - \frac{\left|Y\left(0 \cdot \frac{rad}{sec}\right)\right|}{\sqrt{2}}, \omega \right] \right| \qquad \omega_c = 150.958 \cdot \frac{rad}{sec}$$

Check: $MAG(\omega_c) = -3.01$

which corresponds, as we expect, to a 3 decibel drop. The bandwidth is equal to the cutoff frequency, in this case, since the first cutoff frequency is zero.

The time-domain output of the system is

$\omega_d := \omega_n \cdot \sqrt{1 - \zeta^2}$ $envelope(t) := \frac{\omega_n \cdot e^{-\zeta \cdot \omega_n \cdot t}}{\omega_d}$

$$y(t) := 1 - envelope(t) \cdot \sin\left(\omega_d \cdot t + \operatorname{atan}\left(\frac{\omega_d}{\zeta \cdot \omega_n}\right)\right)$$

In the time-domain, examine the overshoot and the dominant time constant. The dominant time constant is given by inspection of the solution, from which you can see that the transient response is the decaying exponential. The time constant is the multiplier in this exponential, described as the function **envelope(t)** above.

$\tau := \frac{1}{\zeta \cdot \omega_n}$ $\tau = 0.05 \cdot sec$

The overshoot, as defined in **Chapter 10**, is the maximum difference between the transient and steady state solutions for a unit step input. We can find this value using derivatives and the **root** function, as above:

$D(t) := \frac{d}{dt} y(t)$ Guess: $t := \frac{\pi}{\omega_d}$

Find the time at which the derivative is zero.

$t_{OS} := root(D(t), t)$ $t_{OS} = 0.032 \cdot sec$

The value at this point is

$value := y(t_{OS})$ $value = 1.527$

The steady-state value is approximately the value after 5 time constants:

$F := y(5 \cdot \tau)$ $F = 0.995$

So the overshoot is

$overshoot := F - value$

Now plot both the time and the frequency response, and display the various specifications on the graphs with markers.

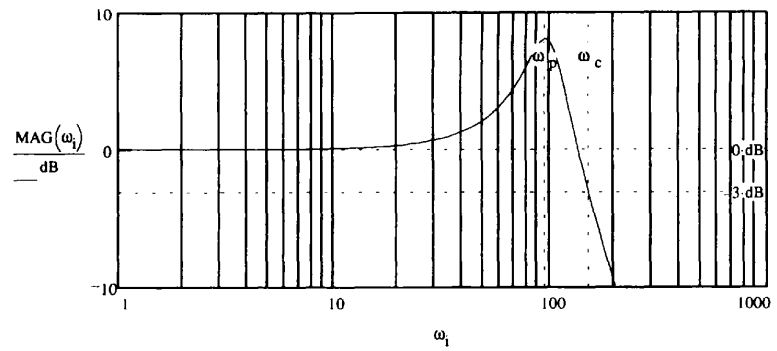
Create a time scale: $t := 0 \cdot sec, .1 \cdot \tau .. 4 \cdot \tau$

To evenly space points on a logarithmic scale, use the following definitions to create the frequency range.

number of points: $N := 100$ $i := 0 .. N - 1$

step size: $r := \log\left(\frac{.01 \cdot \omega_n}{2 \cdot \omega_n}\right) \cdot \frac{1}{N}$

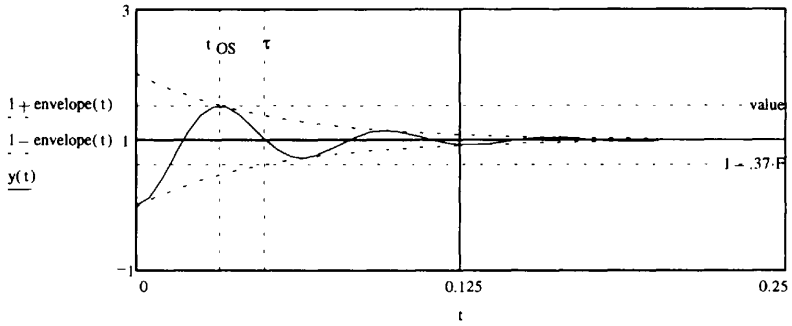
range variable: $\omega_i := 2 \cdot \omega_n \cdot 10^{i \cdot r}$



Change these:

$\omega_n := 100 \cdot \frac{rad}{sec}$

$\zeta := .2$



Experiment with the values of the natural frequency and the damping ratio defined next to the graphs. As always, the accuracy of the answers you get will depend somewhat on the guess value you choose for the root-finding routines. An effort has been made to build a guess which works for most values, but be careful to check that answers make physical sense. You may need to adjust the guess in some extreme cases.

What happens to the various specifications as the damping ratio changes? What about the natural frequency? What does this tell you in terms of system design?

Nyquist Analysis of Time-Delayed Systems

(Schaum's Feedback and Control Systems, 2nd ed., Supplementary Problem 11.80, p. 296)

Statement Plot the Nyquist diagram for the following for time delayed **GH(s)** shown below

System Parameters

$$GH(s) := \frac{e^{-s}}{s\cdot(s+1)}$$

Solution Parametrize the path in the **s**-plane in four pieces:

Number of points per segment: $n := 500$ $m := 0..n$

Small deviation around pole: $\rho := .2$

Radius of semicircle in the **s**-plane: $R := 100$

Draw a semicircle around the pole on the jω-axis.

$$s_m := \rho\cdot e^{j\cdot\left(\frac{\pi\cdot m}{n}-\frac{\pi}{2}\right)}$$

Draw a line on the jω-axis from small radius ρ to large radius **R**.

$$s_{n+m} := j\cdot\left[\frac{m\cdot(R-\rho)}{n}+\rho\right]$$

Draw a semicircle of radius **R**.

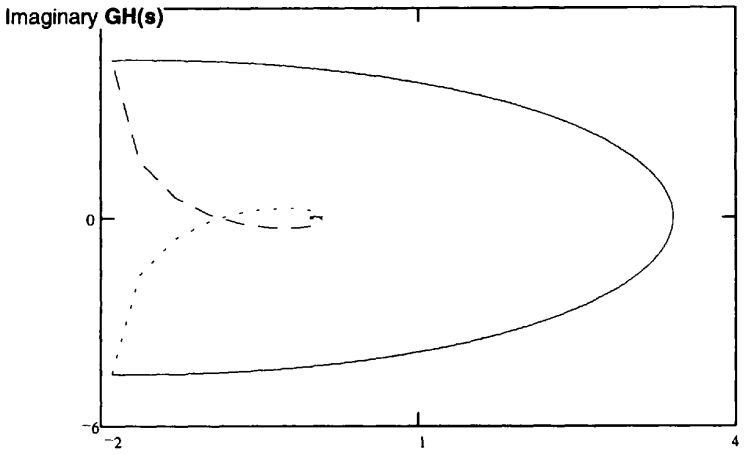
$$s_{2\cdot n+m} := R\cdot e^{j\cdot\left(\frac{-\pi\cdot m}{n}+\frac{\pi}{2}\right)}$$

Draw a line on the jω-axis from large radius **R** to small radius ρ.

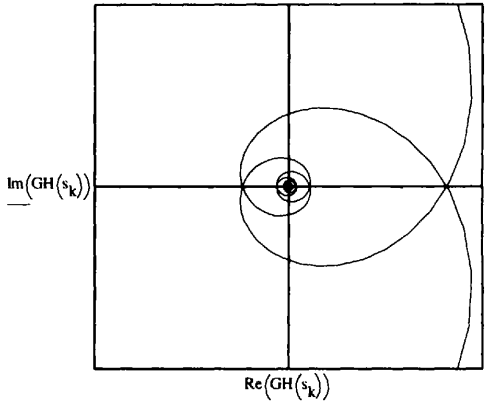
$$s_{3\cdot n+m} := -j\cdot\left[R-\frac{m\cdot(R-\rho)}{n}\right]$$

Close the path and index it: $s_0 := s_{4\cdot n}$ $k := 0..4\cdot n$

Here is the Nyquist diagram for this system. Each part of the path above is mapped with a different line type (solid, dashed, etc.).



Here's an expanded view of the central structure:



The time delay introduces a diminishing spiral to the Nyquist plot of the open-loop transfer function, which spirals in with increasing frequency along the Nyquist path, and back out as the Nyquist path frequency returns to zero. This spiral is superimposed upon the familiar structure you've seen before for a type 1 system in **Chapter 11**.

Gain Factor Compensation Using the Root Locus Method

(Schaum's Feedback and Control Systems, 2nd ed., Solved Problem 14.1, p. 354)

Statement

Determine the value of the gain factor **K** for which the system with the open-loop transfer function **GH(s)** below has closed loop poles with a damping ratio of ζ.

System Parameters

$$GH(s,K) := \frac{K}{s\cdot(s+4)\cdot(s+2)} \qquad \zeta_{req} := 0.5$$

Solution

The closed loop poles will have a damping ratio of ζ when they make an angle of θ degrees with the negative real axis, where θ is defined below.

$$\theta_\zeta := \operatorname{acos}(\zeta_{req})$$

$$\theta_\zeta = 60\cdot\text{deg}$$

We need the value of **K** at which the root-locus crosses the ζ line in the **s**-plane. Do this graphically and analytically in order to verify the answer. Refer to **Chapter 13** to review how to plot root-loci in Mathcad.

$$\frac{C}{R} \equiv \frac{GH(s)}{1+GH(s)} \qquad \text{or,} \qquad \frac{C}{R} \equiv \frac{\frac{K}{s\cdot(s+4)\cdot(s+2)}}{1+\frac{K}{s\cdot(s+4)\cdot(s+2)}}$$

Load the Symbolic Processor from the **Symbolic** menu. Then, select the expanded equation for **C/R** above, and choose **Simplify** from the **Symbolic** menu. This produces the expression for the system characteristic equation (**Chapter 6**) in the denominator:

$$\frac{C}{R} \equiv \frac{K}{\left(s^3+6\cdot s^2+8\cdot s+K\right)}$$

$$s^3+6\cdot s^2+8\cdot s+K_i \equiv 0 \qquad num_{roots} := 3$$

Solving for the roots of this equation, as shown in **Appendix D**,

$$j := 0..num_{roots}-2 \qquad C_{j+1,j} := 1$$

$$coeff(K) := \begin{pmatrix} -K \\ -8 \\ -6 \end{pmatrix} \qquad k := 0..num_{roots}-1$$

$$i := 0..500 \qquad K_i := \frac{i}{10}$$

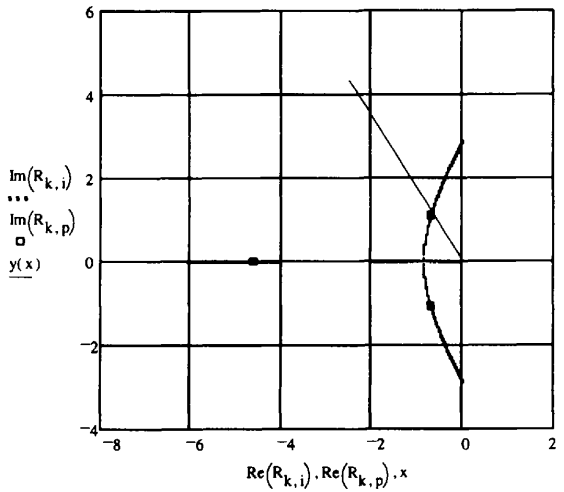
$$R^{\langle i\rangle} := \text{eigenvals}\left(\text{augment}\left(C, coeff\left(K_i\right)\right)\right)$$

The graph of the ζ line is simply a graph of a line with a slope of θ degrees, where the angle was found above. Plot that line by defining **x** and **y(x)** and including them on the root-locus plot.

$$x := -2.5, -2.4..0 \qquad y(x) := \tan\left(-\theta_\zeta\right)\cdot x$$

$$p := 75 \qquad K_p = 7.5$$

Change **p** to see the direction in which the root locus moves with change in gain. This moves the boxes on the trace.



If you change the value of **p** so that one of the boxes moves onto the intersection point of the loci and the damping line, you'll find an approximate value for the desired gain factor, K_p. You can graphically read the value of **s** at which the intersection occurs. Use these values as starting guesses for a **Solve Block**:

$s := -0.5 + 0.8j$ $K := 7.5$

Use the three constraints on the values of **s** and **K**:

Given

$\arg(s) = \pi - \theta_\zeta$ damping ratio constraint

$\arg(GH(s,K)) = -1 \cdot \pi$ angle constraint (**Chapter 13**)

$|GH(s,K)| = 1$ magnitude constraint (**Chapter 13**)

$\begin{pmatrix} s \\ K \end{pmatrix} := \text{Find}(s,K)$ $\begin{pmatrix} s \\ K \end{pmatrix} = \begin{pmatrix} -0.667 + 1.155i \\ 8.296 \end{pmatrix}$

Check the solution:

$\arg(s) = 120 \cdot \text{deg}$ $\pi - \theta_\zeta = 120 \cdot \text{deg}$

$\arg(GH(s,K)) = -1 \cdot \pi$

$|GH(s,K)| = 1$

You should try changing the required value of the damping ratio to see the way the required gain compensation changes. If you do this, remember that you may have to change the guess values for **s** and **K** to get a correct answer from the **Solve Block** above. See **A Mathcad Tutorial** for more information on **Solve Blocks**.

Index