3 STEADY-STATE OSCILLATIONS IN NONLINEAR SYSTEMS

3.0 INTRODUCTION

The preceding chapter introduced the notion of a sinusoidal-input describing function (DF). Some of the implications of this type of *linearization* are discussed there. Here we apply the DF to the study of steady-state oscillations. For DF utilization to be meaningful, certain conditions must be fulfilled by the nonlinearity, and also by the system in which the nonlinearity is present:

1. The nonlinear element is time-invariant.[1]
2. No subharmonics are generated by the nonlinearity in response to a sinusoidal input.
3. The system filters nonlinearity output harmonics to the extent that only a trivial quantity is fed back.

Condition 3 is the so-called *filter hypothesis*. We shall have a great deal to say about it in this chapter, for this heuristically motivated requirement is certainly a basic factor underlying DF success or failure.

Certain periodically varying nonlinearities are also admissible.

Three types of steady-state oscillations are of interest:

1. Forced oscillations
2. Conservative free oscillations
3. Limit cycles

For our present purpose a *forced* oscillation is taken to be a systematic response whose frequency is precisely the forcing signal frequency and whose amplitude depends on the forcing signal amplitude.   Forced oscillations are encountered in the study of frequency response, in which connection we find a useful application for the DF.   The next two oscillation types are behavioral modes of unforced systems.   A *conservative free oscillation* is an *initial condition-dependent* periodic mode associated with nondissipative (conservative) systems.   A *continuous* range of conservative-free-oscillation amplitudes and frequencies may be possible in a given system.   *Limit cycle* denotes an *initial condition-independent* periodic mode occurring in dissipative (nonconservative) systems.   Only a *discrete* set of limit cycle amplitudes and frequencies may exist in a given system.   We shall find the DF a most powerful tool for the study of both types of unforced behavior of nonlinear systems.

3.I DETERMINATION OF LIMIT CYCLES .

The limit cycle phenomenon is deserving of special attention since it is apt to occur in any physical nonlinear system.   A limit cycle can be desirable, for example, by providing the vibration (dither) which minimizes frictional effects in mechanical systems.   In fact, it can be absolutely necessary for proper system performance, as we shall see when studying a certain adaptive control scheme in Chap. 6.   On the other hand, a limit cycle can cause mechanical failure of a control system (destructive limit cycle) or operator discomfort and other undesirable effects, as in an aircraft autopilot.

Although precise knowledge of the waveform of a limit cycle is usually not mandatory, knowledge of the *existence* of a limit cycle, as well as its approximate amplitude and frequency, is a prerequisite to good system design.

THE DF LINEARIZED SYSTEM

Consider a physical nonlinear system as represented by Fig. 3.1-1a.   To effect a limit cycle study, the nonlinear element is characterized by its DF, and the linear element by its frequency response function $L(j\omega)$. $N(A,\omega)$ is used to describe the effect of the nonlinearity on the closed-loop system as we are about to seek the presence of an oscillation which is sinusoidal at
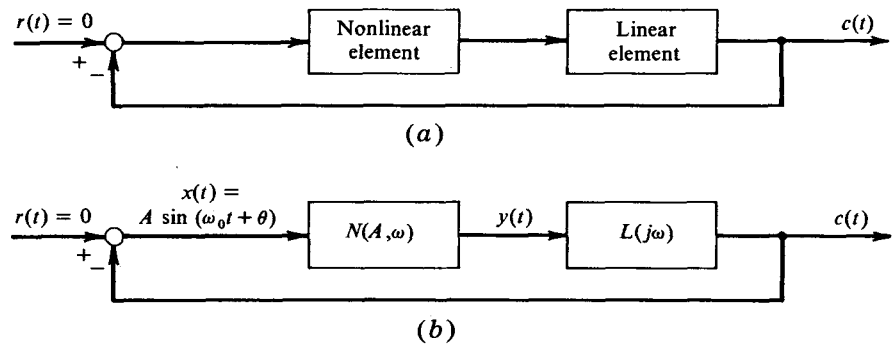
Figure 3.1-1 (a) Single-nonlinearity closed-loop system. (b) Corresponding linearized system for DF limit cycle study.

station x due to the filtering qualities of the loop linear part.   This yields the quasi-linearized system of Fig. 3.1-1b, which, for constant A, is a constant-coefficient linear system.

Linear theory is now applied to the quasi-linearized system, and points of neutral stability are sought.   Any undamped oscillations so arrived at are interpreted as limit cycles in the original nonlinear system.

The equation defining undamped oscillations of the DF linearized system is readily obtained.   Since the input is taken to be zero, the following loop relationship is apparent:

$$X(j\omega) = -L(j\omega)Y(j\omega) \qquad (3.1\text{-}1)$$

Treating the DF as a linearized gain between x and y gives

$$Y(j\omega) = N(A,\omega)X(j\omega) \qquad (3.1\text{-}2)$$

Writing these two equations in matrix notation results in

$$\begin{bmatrix} N(A,\omega) & -1 \\ 1 & L(j\omega) \end{bmatrix} \begin{bmatrix} X(j\omega) \\ Y(j\omega) \end{bmatrix} = 0$$

which possesses a nontrivial solution only if the determinant of the square matrix is zero.   Thus

$$\begin{vmatrix} N(A,\omega) & -1 \\ 1 & L(j\omega) \end{vmatrix} = 0$$

or

$$1 + N(A,\omega)L(j\omega) = 0 \qquad (3.1\text{-}3)$$

Solutions of this equation yield the amplitudes and frequencies of the loop limit cycles.

Before proceeding to both algebraic and graphical solutions of Eq. (3.1-3), we digress to present a different method, which centers attention directly on the characteristic differential equation of the DF linearized system.

ANALYTIC ROUTH-HURWITZ TEST

The Routh-Hurwitz stability criterion for linear systems is well known (cf. Refs. 21, 29, 104). It is a means for determining conditions under which the linear system is neutrally stable. Equation (3.1-3), written as a polynomial in descending powers of s, is the characteristic equation to which this criterion is presently applied. In the case of a third-order system we have, for example,

$$s^3 + a_2 s^2 + a_1 s + a_0 = 0 \tag{3.1-4}$$

The stability boundary is given by

$$a_1 a_2 - a_0 = 0 \tag{3.1-5}$$

provided that a_0, a_1, and a_2 are positive. We now have one relationship in the unknowns A and ω_0. The remaining relationship is a result of the fact that the characteristic equation must contain an undamped second-order factor, the limit cycle. Hence we have

$$(s^2 + \omega_0^2)(s + b) = 0 \tag{3.1-6}$$

Expansion yields

$$s^3 + bs^2 + \omega_0^2 s + b\omega_0^2 = 0$$

whence we observe by comparison with Eq. (3.1-4) that

$$a_1 = \omega_0^2 \tag{3.1-7}$$

This is the second of the required relationships. The limit cycle parameters can now be determined algebraically.

For non-phase-shifting nonlinearities Eq. (3.1-5) is a function of A alone. Once having solved for A, ω_0 follows immediately from Eq. (3.1-7). In order to use this method with phase-shifting nonlinearities, the proportional plus derivative DF formulation [Eq. (2.2-30)] must be employed. Solution of the resulting equations in this case, however, is apt to require graphical assistance since Eq. (3.1-5) is no longer an equation in the single variable A. Simultaneous nonlinear algebraic equations must be dealt with. Table 3.1-1 summarizes the results of this technique as applied to second-, third-, and fourth-order characteristic equations. All a_i must be positive.

The application of this technique for limit cycle determination is demonstrated by the following example.

**TABLE 3.1-1   ROUTH-HURWITZ
LIMIT CYCLE DETERMINATION**

Characteristic equation	Limit cycle equations
$s^2 + a_1 s + a_0 = 0$	$a_1 = 0$ $\omega_0{}^2 = a_0$
$s^3 + a_2 s^2 + a_1 s + a_0 = 0$	$a_1 a_2 - a_0 = 0$ $\omega_0{}^2 = a_1$
$s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0 = 0$	$a_1 a_2 a_3 - a_1{}^2 - a_0 a_3{}^2 = 0$ $\omega_0{}^2 = \dfrac{a_1}{a_3}$

Example 3.1-1   Use the Routh-Hurwitz test to determine all possible limit cycles of the relay control system of Fig. 3.1-2.

The DF for a relay with dead zone, given by Eq. (2.3-14), is repeated below for convenience ($A > \delta$).

$$n_p = \frac{4D}{\pi A}\sqrt{1 - \left(\frac{\delta}{A}\right)^2} \qquad n_q = 0$$

Thus the characteristic equation of interest, $1 + N(A)L(s) = 0$, is representable as

$$s^3 + 2\zeta\omega_n s^2 + \omega_n{}^2 s + \frac{4\omega_n{}^2 KD}{\pi A}\sqrt{1 - \left(\frac{\delta}{A}\right)^2} = 0$$

The coefficients a_i are given by

$$a_2 = 2\zeta\omega_n \qquad a_1 = \omega_n{}^2 \qquad a_0 = \frac{4\omega_n{}^2 KD}{\pi A}\sqrt{1 - \left(\frac{\delta}{A}\right)^2}$$

It follows from the second entry in Table 3.1-1 that the limit cycle frequency is exactly the undamped natural frequency of the second-order part of $L(s)$,

$$\omega_0 = \omega_n$$

and, additionally, that

$$(2\zeta\omega_n)(\omega_n{}^2) - \frac{4\omega_n{}^2 KD}{\pi A_0}\sqrt{1 - \left(\frac{\delta}{A_0}\right)^2} = 0$$
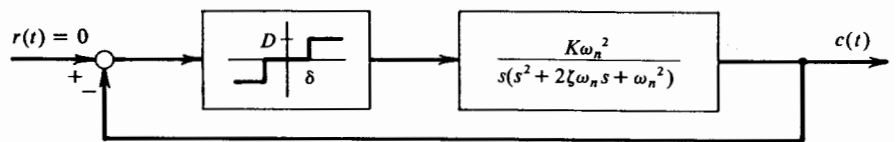

Figure 3.1-2   Example relay control system.

The last of these equations may be solved for the possible limit cycle amplitudes A_0.

$$A_0 = \frac{\sqrt{2}KD}{\pi\zeta\omega_n}\sqrt{1 \pm \sqrt{1 - \left(\frac{\pi\zeta\omega_n\delta}{KD}\right)^2}}$$

It is worthwhile to make several observations regarding these results. First, since the limit cycle amplitude must be a real quantity, we see that no limit cycle is possible for δ in the range

$$\delta > \frac{KD}{\pi\zeta\omega_n} \qquad \text{condition for no limit cycle}$$

Second, when δ is not in the above range, *two* limit cycles are possible, corresponding to the two choices of sign under the radical in the solution found for A_0. We shall see in Sec. 3.2 that of these two limit cycle states, only one is stable.

ALGEBRAIC SOLUTION OF THE LIMIT CYCLE EQUATION

An algebraic solution of Eq. (3.1-3) can be effected directly by expressing $N(A,\omega)$ in its complex form.

$$1 + [n_p(A_0,\omega_0) + jn_q(A_0,\omega_0)]\{\text{Re } [L(j\omega_0)] + j\text{ Im } [L(j\omega_0)]\} = 0$$

This relationship is satisfied when the real and imaginary parts on both sides are equated. The following pair of equations result.

Real: $n_p(A_0,\omega_0) \text{ Re } [L(j\omega_0)] - n_q(A_0,\omega_0) \text{ Im } [L(j\omega_0)] = -1$ (3.1-8a)

Imaginary: $n_p(A_0,\omega_0) \text{ Im } [L(j\omega_0)] + n_q(A_0,\omega_0) \text{ Re } [L(j\omega_0)] = 0$ (3.1-8b)

These two equations in the two unknowns defining the limit cycle are solved simultaneously. When the linear elements are of low order and the DF expressions are relatively simple, analytic solutions for A_0 and ω_0 are readily achieved.

Example 3.1-2 Find the limit cycles displayed by the relay control system of Fig. 3.1-2 by analytically employing Eqs. (3.1-8a) and (3.1-8b).

The DF for a relay with dead-zone nonlinearity, given by Eq. (2.3-14), is repeated here for convenience ($A > \delta$):

$$n_p = \frac{4D}{\pi A}\sqrt{1 - \left(\frac{\delta}{A}\right)^2} \qquad n_q = 0$$

The real and imaginary parts of $L(j\omega)$ are

$$\text{Re } [L(j\omega)] = \frac{-2K\zeta\omega_n^3}{(\omega_n^2 - \omega^2)^2 + (2\zeta\omega\omega_n)^2} \qquad \text{Im } [L(j\omega)] = \frac{-K\omega_n^2(\omega_n^2 - \omega^2)}{\omega[(\omega_n^2 - \omega^2)^2 + (2\zeta\omega\omega_n)^2]}$$

From Eq. (3.1-8b) we conclude that

$$\text{Im } [L(j\omega_0)] = 0$$

This implies that the limit cycle frequency is equal to the natural frequency of the second-order part of $L(j\omega)$.

$$\omega_0 = \omega_n$$

Inserting this value into Eq. (3.1-8a) gives

$$\frac{K}{2\zeta\omega_n} \frac{4D}{\pi A_0} \sqrt{1 - \left(\frac{\delta}{A_0}\right)^2} = 1$$

which has two solutions:

$$A_0 = \frac{\sqrt{2}KD}{\pi\zeta\omega_n} \sqrt{1 \pm \sqrt{1 - \left(\frac{\pi\zeta\omega_n\delta}{KD}\right)^2}}$$

These results are identical with those derived in Example 3.1-1.

An alternative formulation of Eq. (3.1-3) in terms of magnitudes and phase angles rather than real and imaginary parts is

Magnitude: $|N(A_0,\omega_0)|\ |L(j\omega_0)| = 1$ (3.1-9a)

Angle: $\underline{/N(A_0,\omega_0)} + \underline{/L(j\omega_0)} = -\pi - 2n\pi$ (3.1-9b)

for $n = 0, 1, 2, \ldots$. This formulation is clearly equivalent to that of Eqs. (3.1-8a) and (3.1-8b), although one or the other may be more convenient to apply to a specific system.

Example 3.1-3 Determine the limit cycle amplitude and frequency for the nonlinear time-lag differential equation

$$\dot{x}(t) + \delta x(t - T) = \epsilon x^3(t - T)$$

which occurs in the study of cycles in the shipbuilding industry (Ref. 78), nonlinear oscillators with time delay, and certain process control applications.

The block diagram for this system is shown in Fig. 3.1-3. Note that the time-lag behavior is simply represented by the factor e^{-sT}, the Laplace transform of an ideal delay. The equation for a limit cycle is

$$\frac{1}{j\omega_0} e^{-j\omega_0 T}(\delta - \epsilon\tfrac{3}{4}A_0^2) = -1$$

where the DF for a cubing operator, $N(A) = \tfrac{3}{4}A^2$, has been employed. Writing angle and magnitude conditions, we get

$$-\frac{\pi}{2} - \omega_0 T = -\pi - 2n\pi$$

which results in the determination that

$$\omega_0 = (2n + \tfrac{1}{2})\frac{\pi}{T} \qquad \text{for } n = 0, 1, 2, \ldots$$

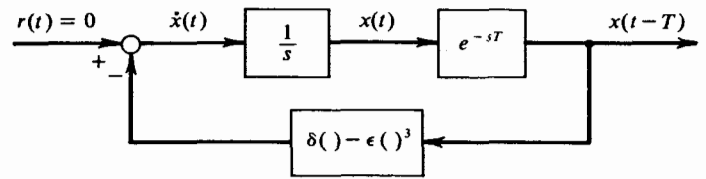and $$\frac{1}{\omega_0} (1)(\delta - \epsilon\tfrac{3}{4}A_0^2) = 1$$

Figure 3.1-3   *Nonlinear time-lag system.*

which implies

$$A_0 = \sqrt{\frac{4}{3\epsilon}\left[\delta - (2n + \tfrac{1}{2})\frac{\pi}{T}\right]}$$

Hence, for values of δ in excess of $\pi/2T$, a limit cycle is predicted. Experimental results substantiate this lower bound. We shall encounter this example once again in Chap. 4. At that point a more detailed discussion of limit cycle behavior is presented.

For systems with linear elements of high order or with relatively complex DFs, and particularly when various nonlinearities are to be studied, the graphical procedure described next is usually found the most expedient means to limit cycle determination.

GRAPHICAL LIMIT CYCLE DETERMINATION

The closed-loop characteristic equation for linear systems with a variable-gain parameter K is usually written in the form

$$KL(j\omega) = -1 \tag{3.1-10}$$

where the *fixed* point $(-1,0)$ takes on the significance of the stability boundary demarcation. Polar plots, Bode plots, gain-phase plots, and (although more indirectly) root-locus plots of familiar linear theory are all graphical mechanisms to focus attention on the relative positions of the system linear elements' locus and the stability point. Equation (3.1-10) may be written, alternatively, as

$$L(j\omega) = -\frac{1}{K} \tag{3.1-11}$$

in which case the stability point is $(-1/K,0)$, a *variable* quantity depending upon the gain K. In this form the effect of the parameter K on the system is singled out. Similarly, a particularly useful form for the characteristic equation of a nonlinear system [Eq. (3.1-3)] is[1]

$$L(j\omega) = -\frac{1}{N(A,\omega)} \tag{3.1-12}$$

---

[1] As pointed out by Gibson (Ref. 29, p. 348), this analogy is not to be viewed as an extension of the Nyquist criterion, it being based on a contour integration around the right half-plane. Equation (3.1-12) has been arrived at by other reasoning.

In concept, $N(A,\omega)$ replaces the linear gain K, although it may be both amplitude- and frequency-dependent.

Whereas satisfaction of the equality of Eq. (3.1-11) implies an undamped second-order oscillation, corresponding satisfaction of Eq. (3.1-12) indicates a limit cycle. By plotting $L(j\omega)$ and $-1/N(A,\omega)$ loci on the same set of
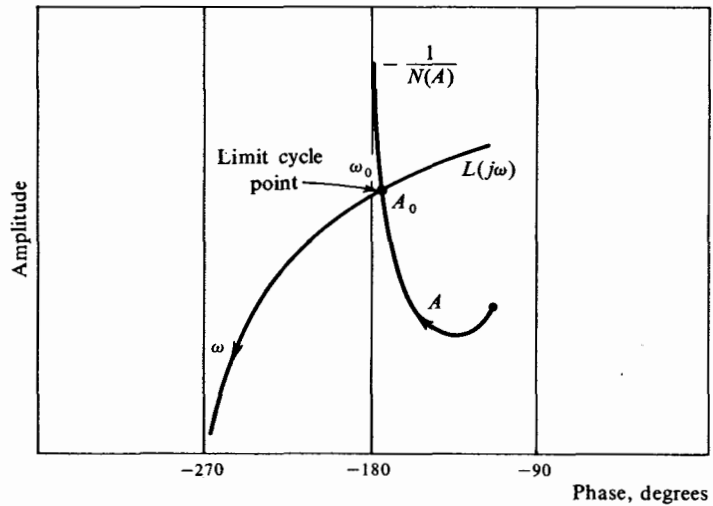
(a)
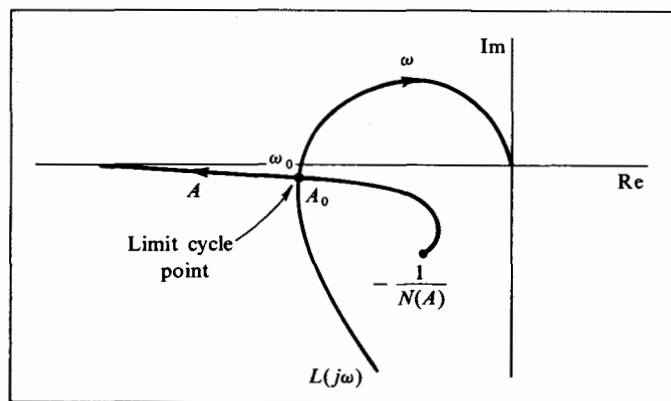
(b)

Figure 3.1-4   *Graphical limit cycle determination. $L(j\omega)$ is third-order, and $N(A)$ is a phase-shifting function (corresponding to a relay with hysteresis).   Amplitude-phase plot (a) and polar plot (b) representations. A_0 and ω_0 are the limit cycle amplitude and frequency, respectively.*

coordinates, possible limit cycle solutions are made to occur as the intersections of two curves. Since one curve represents the loop linear elements and the other represents the nonlinearity, the variation of limit cycle amplitude and frequency with parameter changes of either the linear or nonlinear elements is available with a minimum amount of work by simply altering the appropriate curve.

Limit cycle behavior can be studied in either amplitude-phase (gain-phase), polar, or root-locus coordinates. No limit cycle information exists in any one of these coordinate systems which is not also present in the other two. Since the root-locus construction for phase-shifting nonlinearities is more cumbersome than either of the corresponding gain-phase or polar-plot constructions, it is abandoned here as a working tool.

Typical limit cycle determinations involve constructions such as are shown in Fig. 3.1-4, where third-order linear elements and a phase-shifting DF have been assumed. Arrows indicate the direction of increasing A on the $-1/N(A)$ locus, and increasing ω on the linear elements' frequency locus. The frequency of the indicated limit cycle is read from the calibrated $L(j\omega)$ locus and its amplitude is read from the calibrated $-1/N(A)$ locus.

Whether the polar or amplitude-phase plot is used for this construction is a matter of individual preference. It is the authors' preference to use the amplitude–phase-plane representation, in part because of the general availability of templates for use in linear compensation, and at least in some part because it seems to be a more appealing data presentation.

Example 3.1-4 Graphically identify the two limit cycles found as the solutions to Example 3.1-1.

The nonlinearity is non-phase-shifting; hence $-1/N(A)$ lies along the $-180°$ line on the amplitude-phase plot. From Fig. B.1 we observe that $N(A)$ is not a monotonic function of A; it peaks at exactly $A = \sqrt{2}\delta$. For this reason the amplitude of the $-1/N(A)$ locus starts at $+\infty$, comes down to a minimum at $A = \sqrt{2}\delta$, and returns to $+\infty$, as A increases from zero. In order to illustrate this behavior, the plot of $-1/N(A)$ has been distorted, as shown in Fig. 3.1-5.

Both predicted limit cycles are evident as the intersections of $L(j\omega)$ and $-1/N(A)$. The possibility of no limit cycle is also apparent from this illustration. All that is required is a choice of parameters such that the curves do not intersect at all. As an exercise the reader may formulate this condition from the geometry of Fig. 3.1-5 and compare the result with that calculated in Example 3.1-1.

Limit cycle determination in systems with amplitude- and frequency-dependent DFs is, in principle, quite simple. In practice, the graphical procedure can be tedious. Basically, what one does is to plot several representative members of the family $N(A,\omega)$, each at a *fixed* value of ω. Intersections between $L(j\omega)$ and the various $-1/N(A,\omega)$ loci are examined to determine those for which the frequency *label* of $-1/N(A,\omega)$ and the frequency *calibration* of $L(j\omega)$ match. Such intersections denote the limit
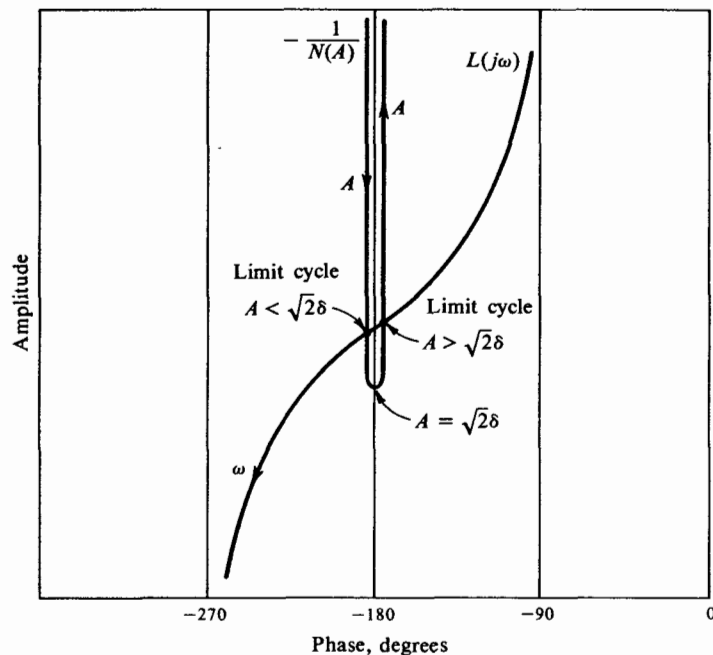
Figure 3.1-5 *Determination of the limit cycles of an example relay control system.*

cycles. Figure 3.1-6 illustrates limit cycle determination in the case where $L(j\omega)$ is third-order and the nonlinear characteristic is a hypothetical relay with frequency-sensitive rectangular hysteresis. At $\omega = \omega_3$ the frequency markings on both curves coincide; hence that is the limit cycle frequency.

We can now determine all possible limit cycles according to DF theory. As pointed out in the text, some of these are stable, others are not.

3.2 LIMIT CYCLE STABILITY

The question of limit cycle stability is posed in terms of the behavior of a theoretical limit cycle state following amplitude and/or frequency perturbations. If the limit cycle returns to its original equilibrium state it will be called *stable*, whereas if either its amplitude or frequency grows or decays until another equilibrium state is reached, it will be called *unstable*. The possibility of theoretically unbounded amplitude or frequency growth is included in the definition of instability by considering the state at infinity to be an equilibrium state.
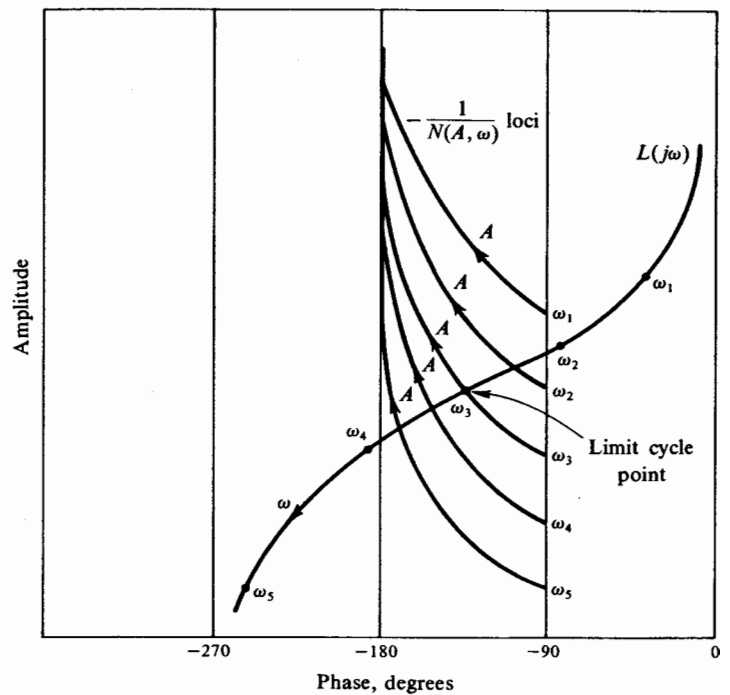
Figure 3.1-6 *Graphical limit cycle determination for a frequency-sensitive nonlinearity. $L(j\omega)$ is third-order, and $N(A,\omega)$ is phase-shifting. Limit cycle frequency is ω_3 radians/time.*

In terms of the behavior of small perturbations, the question of limit cycle stability is a question about a transient oscillation. In that sense it is most properly treated in Chap. 4. Nevertheless, by making what we shall call "quasi-static" arguments,[1] it will be possible to account for limit cycle stability in the great majority of cases. Analytical and graphical quasi-static stability tests proceed easily within the framework already established.

ANALYTICAL STABILITY DETERMINATION

The stability of a limit cycle under small perturbations about an equilibrium state is readily studied in terms of the DF linearized system characterized by Eq. (3.1-3). We follow Loeb's discussion (Ref. 61) of a method originally due to Cahen (Ref. 13). Identical results have been obtained from a geometrical argument by Popov (Ref. 88, p. 586).

---

[1] That is, arguments ignoring terms involving rate of change of frequency and rate of change of damping. See Chap. 4.

Calling A_0 and ω_0 the amplitude and frequency of an equilibrium limit cycle state, respectively, we have seen that the following identity holds:

$$1 + N(A_0,\omega_0)L(j\omega_0) = 0 \qquad (3.2\text{-}1)$$

The explicit complex form of this equation can be obtained by expressing the quantities $N(A_0,\omega_0)$ and $L(j\omega_0)$ in terms of their real and imaginary parts, resulting in

$$U(A_0,\omega_0) + jV(A_0,\omega_0) = 0 \qquad (3.2\text{-}2)$$

where the definitions of U and V are evident.   Let us now allow small perturbations in the limit cycle amplitude, rate of change of amplitude, and frequency by introducing the following changes in Eq. (3.2-2):

$$\begin{aligned} A_0 &\to A_0 + \Delta A \\ \omega_0 &\to \omega_0 + \Delta\omega + j\,\Delta\sigma \end{aligned} \qquad (3.2\text{-}3)$$

The perturbation in the rate of change of amplitude has been associated with the frequency term, a device which becomes clear upon thinking of the limit cycle in the form $A_0 \exp(j\omega_0 t)$.   That is, $\Delta\sigma = -\dot{A}/A$.   Hence we have

$$U(A_0 + \Delta A, \omega_0 + \Delta\omega + j\,\Delta\sigma) + jV(A_0 + \Delta A, \omega_0 + \Delta\omega + j\,\Delta\sigma) = 0$$
$$(3.2\text{-}4)$$

By definition, ΔA, $\Delta\omega$, and $\Delta\sigma$ are small quantities.   The Taylor series expansion of Eq. (3.2-4) about the equilibrium state, valid to first-order terms, becomes, after removal of the quiescent terms [Eq. (3.2-2)],

$$\frac{\partial U}{\partial A}\Delta A + \frac{\partial U}{\partial \omega}(\Delta\omega + j\,\Delta\sigma) + j\frac{\partial V}{\partial A}\Delta A + j\frac{\partial V}{\partial \omega}(\Delta\omega + j\,\Delta\sigma) = 0 \quad (3.2\text{-}5)$$

Satisfaction of this equation requires that its real and imaginary parts separately vanish.

$$\begin{aligned} \frac{\partial U}{\partial A}\Delta A - \frac{\partial V}{\partial \omega}\Delta\sigma + \frac{\partial U}{\partial \omega}\Delta\omega &= 0 \\[4pt] \frac{\partial V}{\partial A}\Delta A + \frac{\partial U}{\partial \omega}\Delta\sigma + \frac{\partial V}{\partial \omega}\Delta\omega &= 0 \end{aligned} \qquad (3.2\text{-}6)$$

Eliminating $\Delta\omega$ from this set of equations yields a single relationship which can be put in the form

$$\left(\frac{\partial U}{\partial A}\frac{\partial V}{\partial \omega} - \frac{\partial U}{\partial \omega}\frac{\partial V}{\partial A}\right)\Delta A = \left[\left(\frac{\partial U}{\partial \omega}\right)^2 + \left(\frac{\partial V}{\partial \omega}\right)^2\right]\Delta\sigma \qquad (3.2\text{-}7)$$

For a limit cycle to be stable, a positive increment ΔA must lead to a positive $\Delta\sigma$, and similarly, a negative ΔA must cause a negative $\Delta\sigma$.   In other words,

the sign of $\Delta\sigma/\Delta A$ must always be positive for a stable limit cycle condition to exist. Hence, for the proposed limit cycle equilibrium condition to be stable, it is necessary that

$$\frac{\partial U}{\partial A}\frac{\partial V}{\partial \omega} - \frac{\partial U}{\partial \omega}\frac{\partial V}{\partial A} > 0 \qquad (3.2\text{-}8)$$

Example 3.2-1 Determine the stability of each of the limit cycles found in Example 3.1-1.

The equation to which we shall apply the analytic stability test is, after simplification $(s = j\omega)$,

$$-2\zeta\omega_n\omega^2 + \frac{4DK\omega_n^2}{\pi A}\sqrt{1 - \left(\frac{\delta}{A}\right)^2} + j(\omega\omega_n^2 - \omega^3) = 0$$

the real and imaginary parts of which are U and V, respectively. Taking the required four partial derivatives and evaluating them at (A_0, ω_0) results in

$$\left.\frac{\partial U}{\partial A}\right|_{A_0,\omega_0} = -\frac{4DK\omega_n^2}{\pi A_0^2}\left[\frac{1 - 2(\delta/A_0)^2}{\sqrt{1 - (\delta/A_0)^2}}\right] \qquad \left.\frac{\partial V}{\partial A}\right|_{A_0,\omega_0} = 0$$

$$\left.\frac{\partial U}{\partial \omega}\right|_{A_0,\omega_0} = -4\zeta\omega_n^2 \qquad \left.\frac{\partial V}{\partial \omega}\right|_{A_0,\omega_0} = -2\omega_n^2$$

Inserting these values into Eq. (3.2-8) immediately yields, as the requirement for a stable oscillation, the condition that

$$1 - 2\left(\frac{\delta}{A_0}\right)^2 > 0$$

or equivalently,

$$A_0 > \sqrt{2}\delta$$

We have already ascertained in Example 3.1-4 that, of the two limit cycles which are predicted, one corresponds to $A_0 > \sqrt{2}\delta$ and one to $A_0 < \sqrt{2}\delta$. From the above analysis it is clear that only the larger-amplitude limit cycle is stable.

GRAPHICAL STABILITY DETERMINATION

The stability of a limit cycle state may be assessed graphically, under the same qualifications which apply to the analytic tests. The argument is borrowed from linear-system stability theory, namely, if the limit cycle amplitude perturbation is positive, we require a stable system configuration in which energy is dissipated until the amplitude decays to its unperturbed value; whereas if the amplitude perturbation is negative, we require an unstable system configuration in which energy is absorbed until the amplitude grows to its unperturbed value. This behavior guarantees a locally stable limit cycle state.

Figure 3.2-1 illustrates amplitude-phase, polar, and root-locus plots for a nonlinear system with a single-valued saturating nonlinearity [$N(A)$ is real and decreases with increasing A]. In the root-locus plot the entire quantity
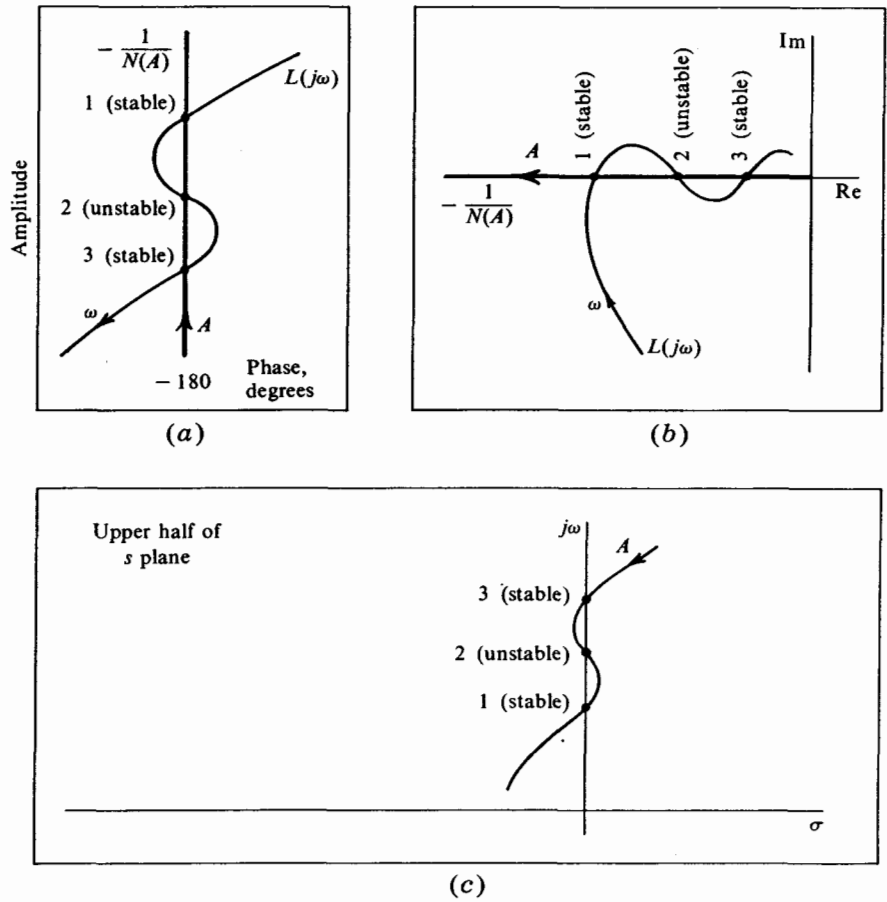
Figure 3.2-1 *Amplitude-phase plot (a), polar plot (b), and root-locus plot (c) of a nonlinear system displaying two stable and one unstable limit cycle modes.*

$N(A)L(s)$ is dealt with [corresponding to $KL(s)$ in the totally linear case]; the arrow on the root-locus branch illustrates pole migration for increasing A. Limit cycles in this construction are determined by intersections of the locus of roots and the $j\omega$ axis. Now, by the argument described above, we find that limit cycle states 1 and 3 are stable (convergent), whereas limit cycle state 2 is unstable (divergent). Consider the high-frequency intersection, state 3, for example. A positive perturbation in limit cycle amplitude $(A_3 \rightarrow A_3 + \Delta A)$ places the stability point $[-1/N(A_3 + \Delta A), 0]$ to the left of the equilibrium state in the polar-plane representation and up in the amplitude–phase-plane representation, and puts a closed-loop root in the left half

of the s plane, each of which would correspond to a stable configuration in the totally linear case. Conversely, a negative perturbation in A_3 leads to an unstable system configuration. Hence, for small perturbations, the resultant limit cycle converges to its original state with increasing time; it is locally *stable* (convergent).

Example 3.2-2 Determine the stability of each of the limit cycles found in Example 3.1-1.
Inspecting pertinent Fig. 3.1-5, it is clear from the argument above that only the larger-amplitude limit cycle oscillation is stable.

Returning to the system of Fig. 3.2-1, we note that if this system is turned on with *zero* initial condition at $c(t)$, the amplitude will grow until state 3 is reached, at which time a stable limit cycle will exist. For positive amplitude perturbations of sufficient magnitude so that state 2 is reached and exceeded, however, the amplitude will eventually converge to state 1 rather than return to state 3. This is a result of the divergent nature of equilibrium state 2.

Physically, limit cycle amplitude perturbations can result from command or disturbance input transients as well as system parameter perturbations. Thus a large transient input to the above system in limit cycle state 3 can cause it to fall into limit cycle state 1. Similarly, a transient decrease in the linear elements leading phase contribution[1] in the frequency band from ω_2 to ω_3, sufficiently large tc temporarily cause only *one* intersection near the frequency ω_1, will also result in the limit cycle falling from state 3 into state 1, provided that the parameter transient is of sufficient duration. This is an *irreversible* situation, since termination of the parameter transient with the system in limit cycle state 1 will not return it to state 3. Clearly, other situations can be proposed which are *reversible;* that is, removal of the parameter transient restores the original system state. A rigorous study of such situations can be found in Bautin (Ref. 6).

3.3 FREQUENCY RESPONSE OF NON-LIMIT-CYCLING NONLINEAR SYSTEMS

Although the frequency response data of a nonlinear system cannot be inverted as in the linear case to obtain the exact time response to any given input signal (linear superposition is invalid), for weakly nonlinear systems this process will yield useful approximate results. Further, despite the fact that control systems do not generally receive purely sinusoidal inputs, for certain specific cases sinusoidal response data may be quite significant. These

[1] Caused by either an increase of lagging phase or a decrease of leading phase in the constituents of $L(s)$.

include a periodic, although nonsinusoidal, input and a nonlinear subsystem within the control loop of a larger system which is either lightly damped (i.e., whose transient response is highly oscillatory) or which displays a limit cycle. Thus, for one reason or another, the sinusoidal response description of a nonlinear system is desirable. It will be shown that the DF may be used to represent nonlinearity behavior in a harmonically excited system, leading to a frequency response solution.

Frequency response as calculated with the DF has significance only if the system under consideration is not in a limit cycle state. It turns out that certain non-limit-cycling systems break into a limit cycle oscillation when forced by a sinusoidal input; limit cycles in certain other systems are quenched by the introduction of a sinusoidal input. It is therefore necessary to examine the state of a sinusoidally forced system to determine whether a limit cycle can exist. This requires the use of a two-sinusoid-input describing function, as we shall see in Chap. 5. Similarly, testing the stability of a DF-derived frequency response solution requires methods discussed in Chap. 5. For the
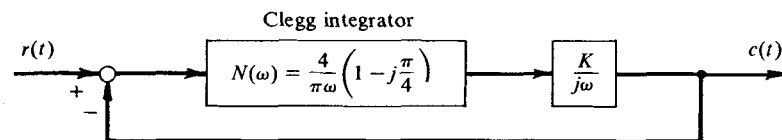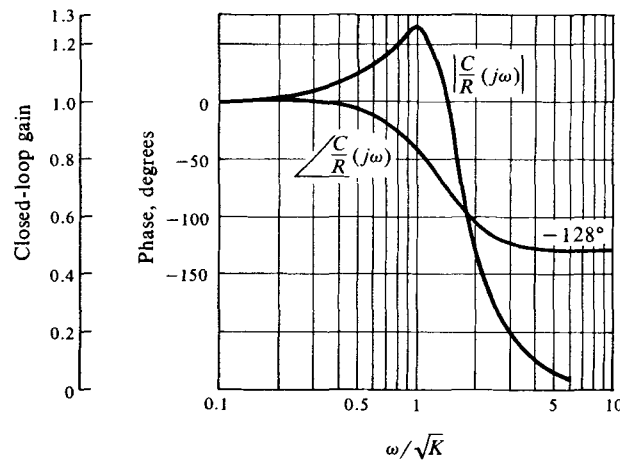
Figure 3.3-1 (a) *System containing a Clegg integrator.* (b) *Its closed-loop frequency response.* (*Levinson, Ref. 60, pt. 4, p. 132.*)

moment, however, we assume DF study of frequency response to be valid wherever applied.

Example 3.3-1   Determine the frequency response of the nonlinear system containing a Clegg integrator, shown in Fig. 3.3-1a.

Since the DF for the Clegg integrator depends on frequency alone [cf. Eq. (2.4-15)], one proceeds as in the linear case to arrive directly at a closed-form solution for $C/R(j\omega)$. Thus

$$\frac{C}{R}(j\omega) = \frac{N(\omega)K/j\omega}{1 + N(\omega)K/j\omega}$$

$$= \frac{1 + j(4/\pi)}{1 - (\omega^2/K) + j(4/\pi)}$$

$$= \frac{\sqrt{1 + (4/\pi)^2}}{\sqrt{(1 - (\omega^2/K))^2 + (4/\pi)^2}} \exp\left[-j \tan^{-1}\left(\frac{(4\omega^2/\pi K)}{1 - (\omega^2/K) + (4/\pi)^2}\right)\right]$$

The magnitude and phase angle determined above are plotted in Fig. 3.3-1b. Note the lower closed-loop phase than would be associated with the corresponding purely linear system.

In general, the DF is a function of both A and ω. It is the dependence upon A which forces the use of more complicated means for determination of frequency response. Referring to Fig. 3.3-2, let $r(t)$ represent the simple harmonic excitation

$$r(t) = M_r \sin \omega t \qquad (3.3-1)$$

Assuming $x(t)$ to be a sinusoid of amplitude A and frequency ω, the transfer function to $c(t)$ becomes

$$\frac{C}{R}(j\omega,A) = \frac{L_1(j\omega)N(A,\omega)L_2(j\omega)}{1 + L_1(j\omega)N(A,\omega)L_2(j\omega)L_3(j\omega)} \qquad (3.3-2)$$

where the nonlinearity is represented by its DF. This is the basis for a frequency response derivation. In this notation, applicable to all system *variables*, capitals are used to designate frequency-domain-transformed quantities of otherwise lowercase-designated time functions. Continuing, we observe that the transfer function relating C to R is dependent upon A
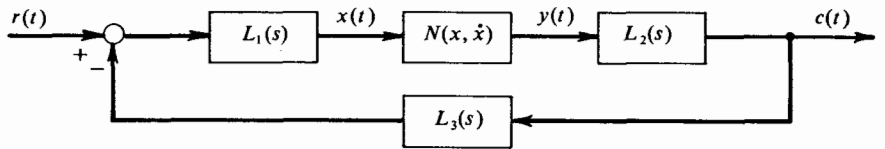

Figure 3.3-2   General nonlinear control system considered.

and ω, the descriptors of an as yet unknown X.   One may negotiate this circular difficulty by working with the related transfer function

$$\frac{X}{R}(j\omega,A) = \frac{L_1(j\omega)}{1 + L_1(j\omega)N(A,\omega)L_2(j\omega)L_3(j\omega)} \qquad (3.3\text{-}3)$$

and considering A and ω as the independent variables (Ref. 68).   At any given input frequency ω, let us assume a value of A.   This procedure, which specifies all of X but its phase, also defines the ratio X/R by virtue of Eq. (3.3-3).   Working with the magnitude of X/R, the value of M_r corresponding to A is immediately determined as

$$M_r = \frac{A}{\left|\dfrac{X}{R}(j\omega,A)\right|} \qquad (3.3\text{-}4)$$

The value of A corresponding to a particular combination of M_r and ω can be found routinely by inspecting accumulated data derived as above. At this point the transfer function C/R [Eq. (3.3-2)] may be determined (for a particular M_r), thus defining the frequency response of the system.   It is clear that all frequency response plots such as either $|C/R|$ versus ω or $\underline{/C/R}$ versus ω must carry the value of M_r for which the plots were derived.

In order to generate a complete set of data relating C to R for the range of expected input amplitudes and frequencies, one must use all the data that relate $N(A,\omega)$ to R in the repeated solutions of Eq. (3.3-3).   A systematized procedure making substantial use of available graphical aids for transfer function evaluation renders this approach worthwhile (Ref. 44).

A CONSTRUCTION FOR FREQUENCY RESPONSE

For circumstances under which it is desirable to study directly the effects of a variety of nonlinearities on the closed-loop frequency response, linear elements remaining fixed, a different solution to the frequency response problem is available.   Let us make the following definitions:

$$\begin{aligned} L_1 &= \rho_1 e^{j\theta_1} \\ L_1 L_2 L_3 &= \rho_2 e^{j\theta_2} \\ N &= \rho_N e^{j\theta_N} \end{aligned} \qquad (3.3\text{-}5)$$

Functional dependence of ρ_1, ρ_2, θ_1, and θ_2 upon ω, and of ρ_N and θ_N upon both ω and A, are omitted for convenience.   Considering the absolute magnitude of each side of Eq. (3.3-3), we may write

$$\begin{aligned} \frac{A}{M_r} &= \left|\frac{\rho_1 \exp(j\theta_1)}{1 + \rho_2\rho_N \exp[j(\theta_2 + \theta_N)]}\right| \\ &= \frac{\rho_1}{\sqrt{1 + 2\rho_2\rho_N \cos(\theta_2 + \theta_N) + (\rho_2\rho_N)^2}} \end{aligned} \qquad (3.3\text{-}6)$$

which can be put in an alternative form,

$$A\rho_N = -\frac{A\cos(\theta_2 + \theta_N)}{\rho_2} \pm \frac{1}{\rho_2} \sqrt{\rho_1{}^2 M_r{}^2 - A^2 \sin^2(\theta_2 + \theta_N)}$$

$$(3.3\text{-}7)$$

This equation may be used to get a graphical solution for the frequency response.   Both sides can be separately plotted as functions of A, ω remaining fixed.   Intersections of the plotted curves represent solutions to Eq. (3.3-3).

For systems with nonlinearities whose DFs are non-phase-shifting, $\theta_N = 0$, and Eq. (3.3-7) reduces to

$$A\rho_N = -\frac{A\cos\theta_2}{\rho_2} \pm \frac{1}{\rho_2} \sqrt{\rho_1{}^2 M_r{}^2 - A^2 \sin^2\theta_2} \qquad (3.3\text{-}8)$$

which is relatively simple to investigate.[1]   To begin with, the right-hand side of this equation is not a function of the DF.   Hence changing the non-linearity does not alter this part of the construction.   For a fixed value of frequency, A is the only independent variable of this equation.   It is readily established that a plot of the right-hand side versus A is an ellipse.

To facilitate the graphical procedure, the key points defining this ellipse are shown in Fig. 3.3-3.   Ordinate and abscissa intercepts are indicated. Coordinates of the maximum ordinate and abscissa excursions serve to locate approximately the tip of the ellipse.   Consider, for a moment, that in the DF study of frequency response, the system involved is required not to limit-cycle.   Where single-valued nonlinearities are present, this restriction implies that θ_2 lies in the range $-\pi < \theta_2 < 0$ (cf. Sec. 3.1).   Since $\csc\theta_2$ is negative throughout this range, the association of $-\csc\theta_2$ with coordinates of the tip of the ellipse in Fig. 3.3-3 is explained.   Because the $\cot\theta_2$ term also appears, we see that for the semimajor axis of the ellipse to actually fall in the first quadrant (as shown), it is required that θ_2 lie in the range $-\pi < \theta_2 < -\pi/2$.

Example 3.3-2   Qualitatively determine the character of the harmonically forced response of the system illustrated in Fig. 3.3-4a.

First we observe that the DF for the system nonlinearity is non-phase-shifting.   For this system we have

$$L_1(s) = L_3(s) = 1 \qquad \text{and} \qquad L_2(s) = \frac{K}{s(s+1)}$$

which imply

$$\rho_1 = 1 \qquad\qquad \theta_1 = 0$$

$$\rho_2 = \frac{K}{\omega\sqrt{1+\omega^2}} \qquad \theta_2 = -\frac{\pi}{2} - \tan^{-1}\omega$$

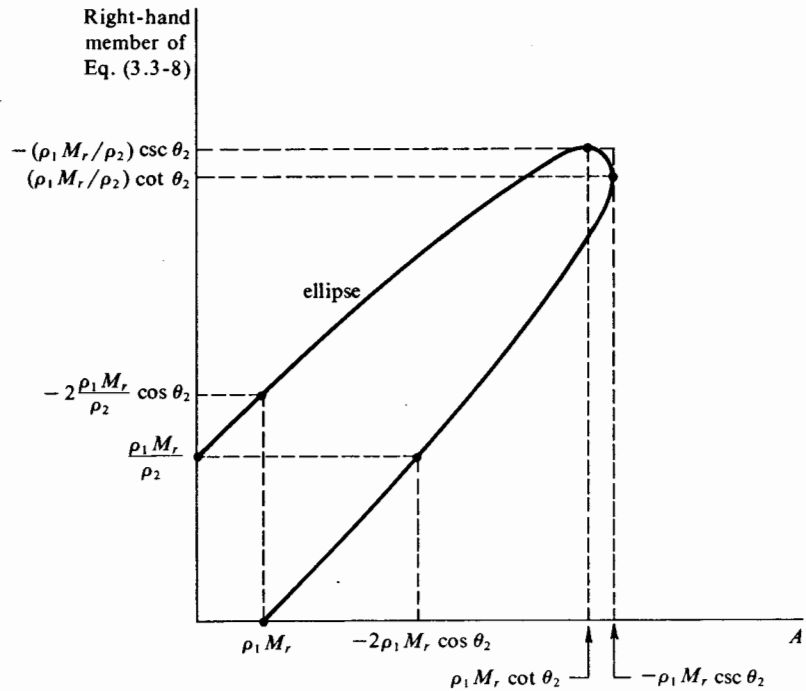[1] We follow Levinson's original work (Ref. 58) in this presentation.

Figure 3.3-3   *A frequency response construction aid, M_r and ω constant.*

Utilizing Fig. 3.3-3, a group of ellipses corresponding to various input frequencies can be drawn. The result of this construction for a single value of M_r is shown in Fig. 3.3-4b. Since $\rho_1 = 1$, all ellipses pass through the point $(M_r, 0)$.

The quantity $A\rho_N$ is simply the amplitude of the nonlinearity output fundamental. In terms of increasing A, the plot of $A\rho_N$ is like a gain-changing function. This can be shown exactly by referring to the DF calculation for the gain-changing nonlinearity available in Chap. 2, or by noting the limiting results

$$A\rho_N \rightarrow m_1 A \qquad \text{as } A \rightarrow 0$$

$$A\rho_N \rightarrow m_2 A \qquad \text{as } A \rightarrow \infty$$

from which a rough plot of $A\rho_N(A)$ can be made. Generally speaking, the shape of $A\rho_N$ closely resembles the nonlinear characteristic itself, with all sharp corners rounded off.

The magnitude of ρ_N can now be found on a point-by-point basis. For example, at $\omega = \omega_5$, it follows that $A = A_5$, and hence that $\rho_N(A_5)$ is known. For certain values of ω, this point-by-point process is unique. Other values of ω result in multiple intersections in Fig. 3.3-4b, indicating a multivalued solution. Utilizing all values of ρ_N found graphically, amplitude and phase plots of $C/R(j\omega)$ can gradually be built up. Figure 3.3-5 illustrates such a solution for the particular M_r of Fig. 3.3-4b.
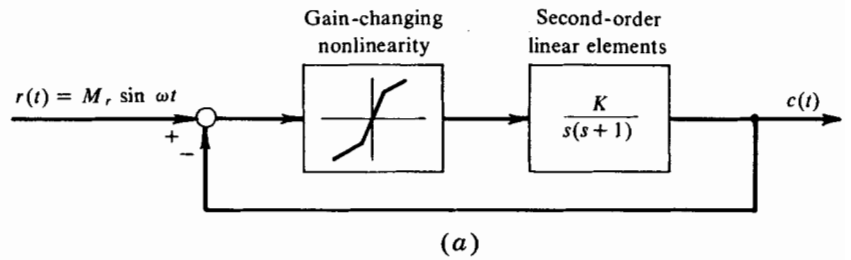
(a)

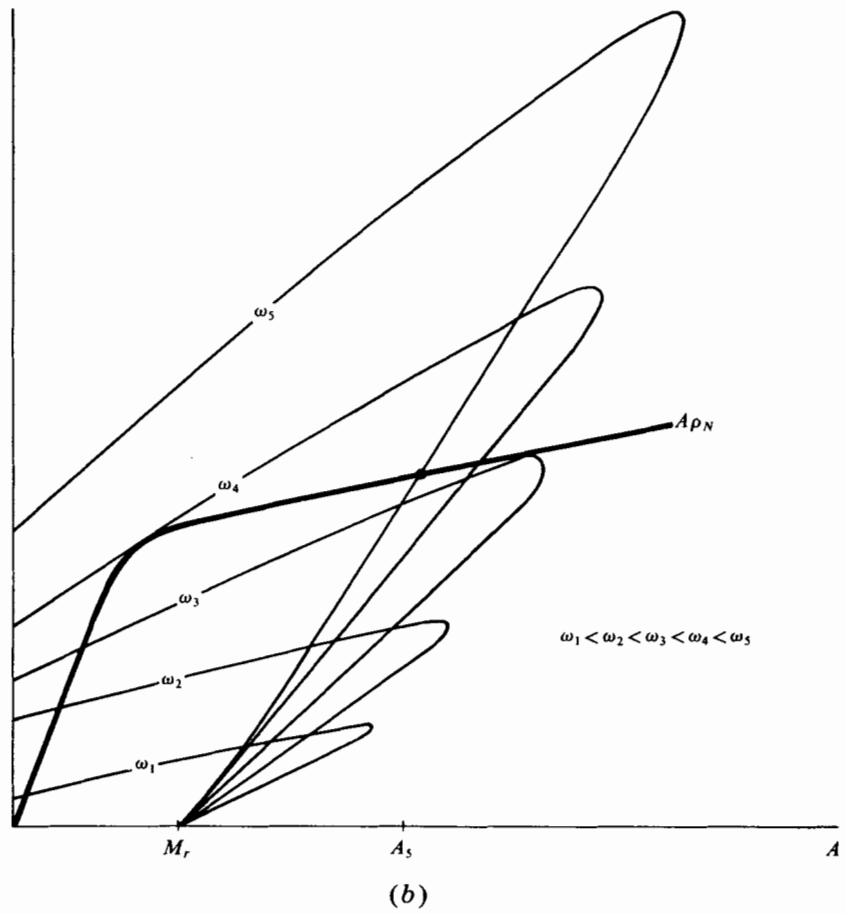**Figure 3.3-4a**   *Example system for frequency response study.*

(b)

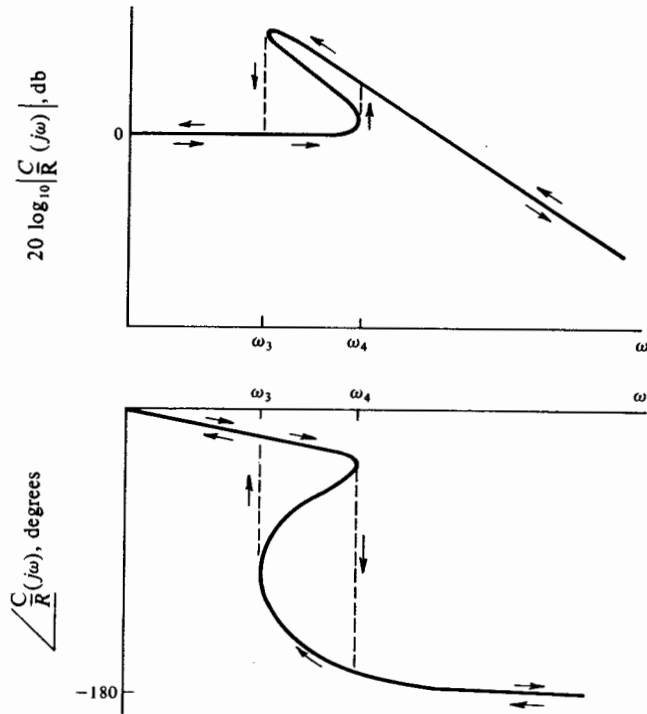**Figure 3.3-4b**   *Frequency response construction for example system.*

Figure 3.3-5 *Closed-loop frequency response curves for a particular M_r. Amplitude and phase jumps occur.*

The closed-loop frequency response curves of the example system appear to exhibit a hysteretic behavior. They indicate discontinuities in the amplitude and phase response for an input frequency either increasing and surpassing ω_4 or decreasing and falling below ω_3. This form of discontinuous closed-loop frequency response has been appropriately labeled *jump resonance*. It is unlike any response provided by a linear system, and is due to the gain-changing character of a system nonlinearity. As a result of the jumps for increasing and decreasing frequency, a portion of the triple-valued amplitude and phase response curves is apparently never traversed. This portion of the curves may therefore be considered as representing unstable steady-state solutions. In fact, this portion cannot be observed experimentally. Proof of such an instability has been demonstrated by Stoker (Ref. 99) and Bonenn (Ref. 10).

Generally speaking, the amplitude resonance peak bends to the left for saturating nonlinear characteristics and to the right for increasing gain characteristics. This is illustrated in Fig. 3.3-6. As implied before, arrows indicate permissible motion along the locus in question.
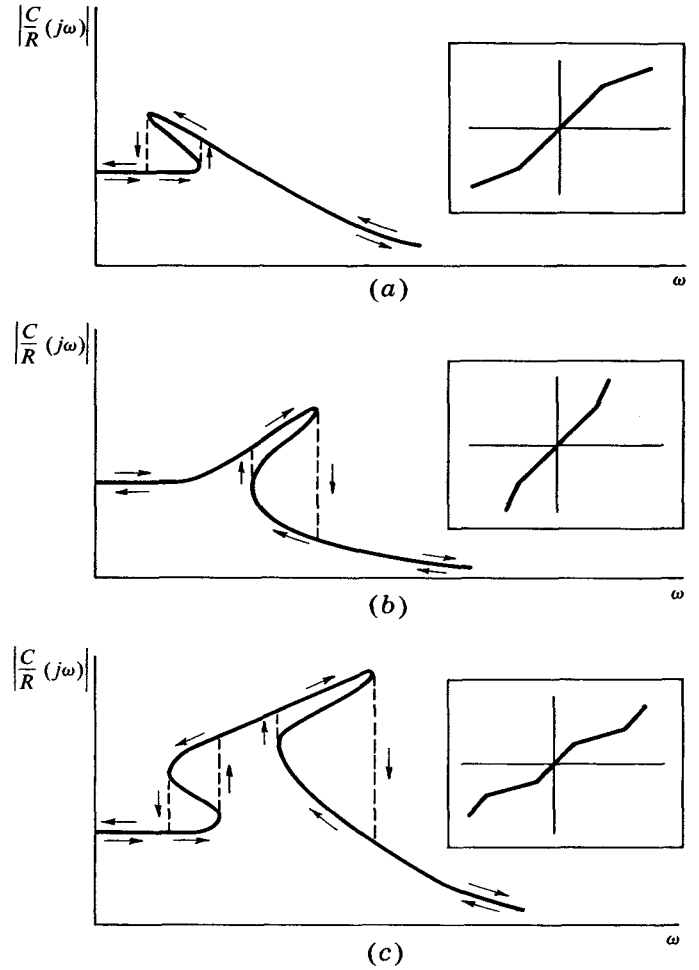
$$\left|\frac{C}{R}(j\omega)\right|$$

(a) ω

$$\left|\frac{C}{R}(j\omega)\right|$$

(b) ω

$$\left|\frac{C}{R}(j\omega)\right|$$

(c) ω

Figure 3.3-6 Decreasing (a), increasing (b), and combination (c) non-linear gain characteristics shown with typical system amplitude jump responses.

Note that for a linear device $A\rho_N$ is directly proportional to A; it follows that a construction similar to that of Fig. 3.3-4b will yield only single intersections with each ellipse. The amplitude resonance peak in this instance is neither to the left nor to the right; *jump resonance cannot occur.*

CRITICAL JUMP RESONANCE CURVE

A curve which encloses a region of the complex plane such that jump resonance can occur if the open-loop frequency response locus of the system linear

part passes into this region is called a *critical jump resonance curve*. It is derived in the following discussion after a method due to Hatanaka (Ref. 43). The nonlinearity is restricted to be a single-valued odd static operator.[1]

Recall that at the jump resonance point $|C/R|$ is discontinuous. It follows that $|X/R|$ is also discontinuous. That is, considering the input amplitude M_r to be a function of the error amplitude A for a fixed angular frequency ω, the same value of M_r corresponds to different values of A. If the nonlinearity DF is differentiable with respect to A and possesses a continuous first derivative, the condition under which jump resonance occurs is thus

$$\left(\frac{\partial M_r}{\partial A}\right)_{\omega=\text{const}} = 0 \qquad \text{at the jump resonance point} \qquad (3.3\text{-}9)$$

This condition prevails at points 1 and 2 of Fig. 3.3-7, which is a typical plot of A versus M_r (at fixed ω) for the system illustrated. Further, corresponding to the region between points 1 and 2, the requirement for unstable forced responses is that the slope of the curve be negative (Ref. 10).

$$\left(\frac{\partial M_r}{\partial A}\right)_{\omega=\text{const}} < 0 \qquad \text{for region in which jump resonance occurs} \qquad (3.3\text{-}10)$$

Using the above results, we can proceed to the derivation of interest. First, note that by use of the definition

$$L(j\omega) = U(\omega) + jV(\omega)$$

one can put the *magnitude* of the system error transfer function

$$\frac{X}{R}(j\omega,A) = \frac{1}{1 + N(A)L(j\omega)}$$

in the form

$$M_r = AN(A)\sqrt{\left[U(\omega) + \frac{1}{N(A)}\right]^2 + [V(\omega)]^2} \qquad (3.3\text{-}11)$$

Taking the partial derivative of M_r in accordance with Eq. (3.3-9) yields

$$\left[U(\omega) + \frac{1}{N(A)}\right]\left[U(\omega) + \frac{1}{N^*(A)}\right] + V(\omega)^2 = 0 \qquad (3.3\text{-}12)$$

where N^* is defined by

$$N^*(A) = N(A) + AN'(A) \qquad (3.3\text{-}13)$$

and the prime indicates differentiation with respect to A. From Eq. (3.3-12) we see that contours in the polar plane for constant values of A at the jump

[1] Another treatment of this problem, including results for the case of a nonlinearity with memory, can be found in Ref. 25.
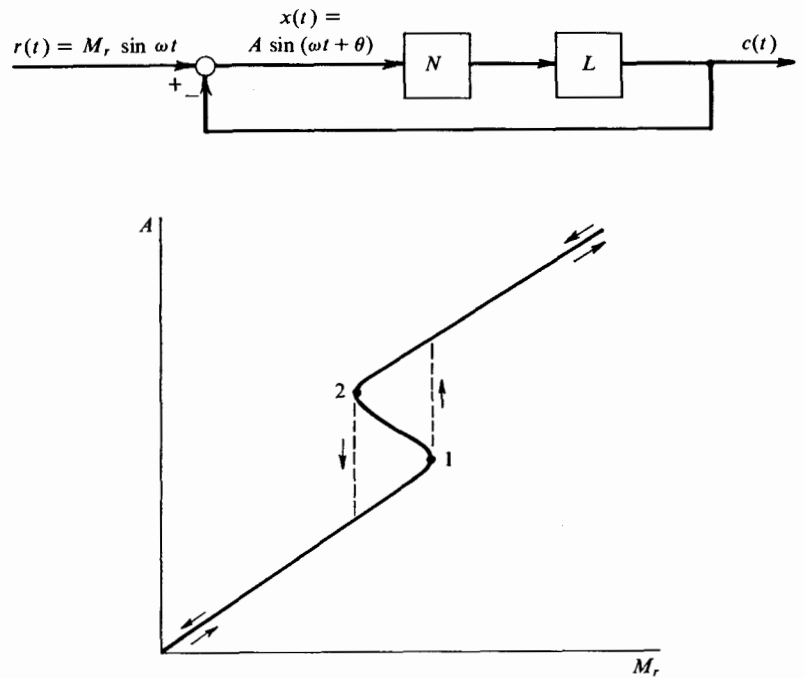
Figure 3.3-7 *Steady-state error amplitude vs. forcing amplitude in a system exhibiting jump resonance.*

resonance point are a family of circles.   Denote this family of contours $f(U,V,A) = 0$.

We now seek the *envelope* of the family of contours $f(U,V,A) = 0$.   It is proved in calculus that such an envelope satisfies the relationships

$$f(U,V,A) = 0 \quad \text{and} \quad \frac{\partial f(U,V,A)}{\partial A} = 0 \tag{3.3-14}$$

Operating on Eq. (3.3-12) in accordance with Eqs. (3.3-14) results in

$$U = -\frac{N'(A)N^*(A) + N(A)N^{*\prime}(A)}{N'(A)[N^*(A)]^2 + N^{*\prime}(A)[N(A)]^2}$$

$$V = \pm \frac{\sqrt{N'(A)N^{*\prime}(A)}\,[N(A) - N^*(A)]}{N'(A)[N^*(A)]^2 + N^{*\prime}(A)[N(A)]^2} \tag{3.3-15}$$

which is the desired critical jump resonance curve specified parametrically in A.

Example 3.3-3   Find the critical jump resonance curve for the nonlinearity $y = x^3$.
From Eqs. (3.3-15) we directly obtain

$$U = -\frac{(\frac{3}{2}A)(\frac{9}{4}A^2) + (\frac{3}{4}A^2)(\frac{9}{2}A)}{(\frac{3}{2}A)(\frac{81}{16}A^4) + (\frac{9}{2}A)(\frac{9}{16}A^4)} = -\frac{2}{3}\frac{1}{A^2}$$

$$V = \pm\frac{\sqrt{(\frac{3}{2}A)(\frac{9}{2}A)(\frac{3}{4}A^2 - \frac{9}{4}A^2)}}{(\frac{3}{2}A)(\frac{81}{16}A^4) + (\frac{9}{2}A)(\frac{9}{16}A^4)} = \pm\frac{2}{3\sqrt{3}}\frac{1}{A^2}$$

Eliminating A yields

$$V = \pm\frac{1}{\sqrt{3}}U$$

where it is understood that U takes only negative values.   This locus is plotted in Fig. 3.3-8.
Crosshatching indicates the interior of the region in which jump resonance occurs since
everywhere within this region Eq.(3.3-10) is satisfied.

The critical jump resonance curves for a gain-changing nonlinearity,
derived as above, are shown in Fig. 3.3-9.   We shall have occasion to recall
these results in Chap. 5, where they are again derived, albeit from quite a
different point of view.

The methods discussed for obtaining the frequency response of a nonlinear
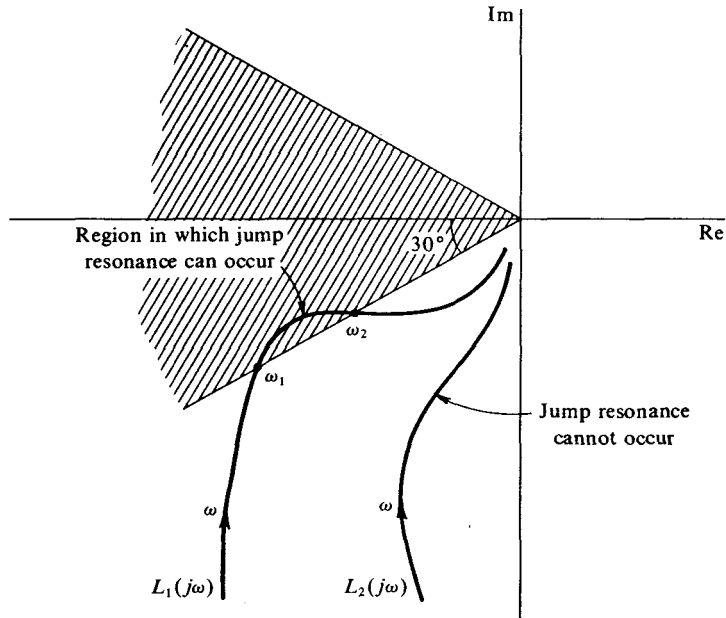system can be applied as well to nonlinear subsystems such as were treated in

Figure 3.3-8   Critical jump resonance curve for cubic nonlinearity.
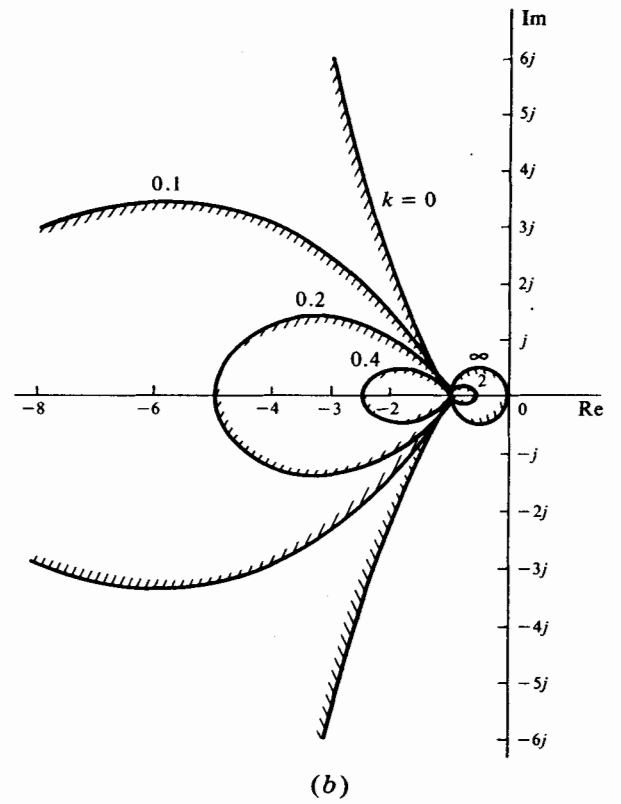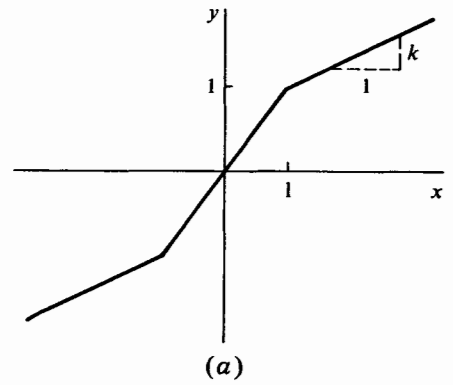
(a)

(b)

Figure 3.3-9 Critical jump resonance curve (b) for gain-changing nonlinearity (a). (Hatanaka, Ref. 43.)

Sec. 2.4. In particular, we now have additional means for DF derivation in the case of implicit dynamical nonlinearities. Before leaving the topic of nonlinear-system frequency response, one additional system is well worth considering, the time-optimal computer control system.

3.4 APPLICATION TO A TIME-OPTIMAL COMPUTER CONTROL SYSTEM

A time-optimal control system is one in which the error and its $n - 1$ derivatives (where n is the order of the controlled process) are simultaneously returned to zero in the shortest possible time. For a $1/s^n$ plant, the system input is characterized by an n-term Taylor series. To demonstrate DF utilization for frequency response investigation, we shall choose a simple $1/s^2$ controlled process in the system configuration of Fig. 3.4-1. An ideal relay is the controlled-process driver, as in all time-optimal systems. This follows from the desire to achieve either *maximum* load acceleration or *maximum* load deceleration in a correspondingly maximum control effort operation. The computer receives all system inputs and outputs, operating on them in such a way as to switch the ideal relay according to the requirements for time-optimal control. Examples of the controlled-process and relay-driver combination depicted include an interplanetary vehicle driven for attitude control about each of two axes by a two-position continuously operating microrocket (Ref. 25), an acceleration-switching hydraulic servovalve (Ref. 38), etc. Sparacino (Ref. 96) has treated this and other second-order processes. The method to be described closely follows his presentation.

The *control law* implemented by the computer must be derived as a prelude to DF evaluation. In order to do this we utilize the fact that for an nth-order system with real poles the error and its $n - 1$ derivatives may be returned to zero time optimally, with at most $n - 1$ switchings. We thus require but one switching. The system error trajectory will therefore be comprised of two parts, the second part necessarily going through the origin of the phase plane (the coordinates of which are error rate versus error).
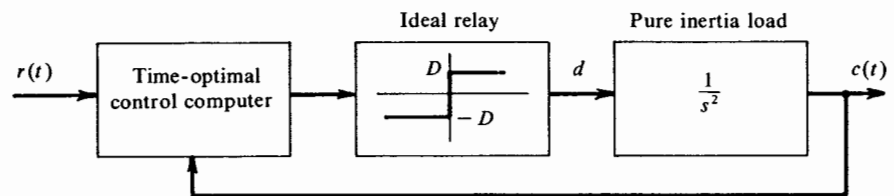

Figure 3.4-1   *Time-optimal control scheme for a second-order controlled process.*

Assuming that the system input may be represented by a two-term Taylor series, the equation of motion for the second-order process at hand during the *last* (second) trajectory may be written as

$$e = e_s + \dot{e}_s(t - t_s) - \frac{d}{2}(t - t_s)^2 \qquad \text{for } t \geqslant t_s \qquad (3.4\text{-}1)$$

where d is the ideal-relay output $(+D$ or $-D)$,

$$e = r - c \qquad (3.4\text{-}2)$$

and e_s is the final value of e for the *initial* trajectory. We now impose the requirements of time-optimal control, namely,

$$e(T) = \dot{e}(T) = 0 \qquad (3.4\text{-}3)$$

where T is the shortest time in which Eq. (3.4-3) can be achieved. Figure 3.4-2 illustrates a typical input, output, and corresponding error trajectory. Solving Eqs. (3.4-1) and (3.4-3) by eliminating time gives

$$e_s = -\frac{\dot{e}_s{}^2}{2d} \qquad (3.4\text{-}4)$$

We must therefore switch when the error satisfies Eq. (3.4-4), for in that case the last trajectory will be perfectly executed, having begun with the proper initial conditions. The control computer therefore chooses the sign of d according to (sgn = signum = "sign of")

$$\operatorname{sgn} d = \operatorname{sgn} \left(e + \frac{\dot{e}^2}{2D} \operatorname{sgn} \dot{e} \right) \qquad (3.4\text{-}5)$$

We are now in a position to derive the DF for the computer and relay, in combination. Since the computer acts upon the quantity $r - c$, the system of Fig. 3.4-1 may be redrawn as in Fig. 3.4-3. $N(e, \dot{e})$ represents the dynamics implied by Eq. (3.4-5). In particular, the output of N is a square wave of amplitude D (first harmonic $4D/\pi$) and phase shift θ_N (see Fig. 3.4-4). Assuming a sinusoidal input

$$e = A \sin \omega t \qquad (3.4\text{-}6)$$

it follows from Eq. (3.4-4) that the switching criterion is

$$A \sin \omega t = -\frac{(A\omega)^2 \cos^2 \omega t}{2d} \qquad (3.4\text{-}7)$$
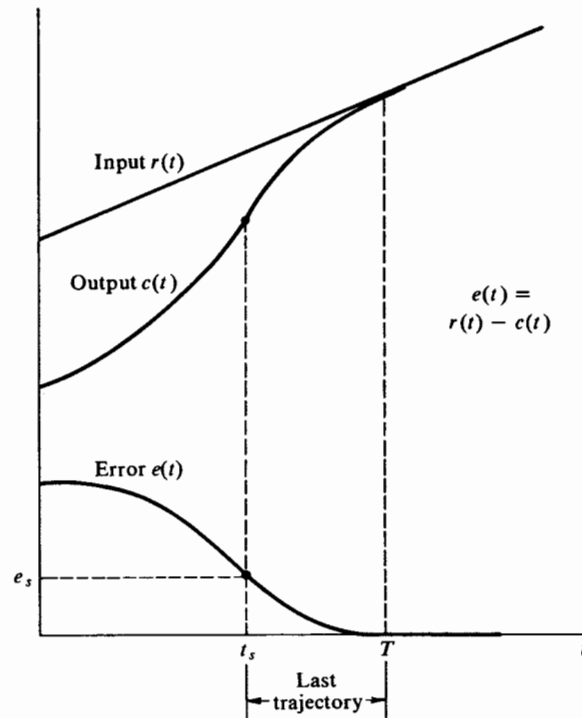
Figure 3.4-2 *Typical time history of the input, output, and error in a time-optimal control system.*

The substitutions $\omega t = -\theta_N$ and $d = D$, indicated by Fig. 3.4-4, yield θ_N in the form

$$\frac{\sin \theta_N}{\cos^2 \theta_N} = \frac{A\omega^2}{2D} \qquad (3.4\text{-}8)$$

or equivalently,

$$\theta_N = \sin^{-1}\left[-\frac{D}{A\omega^2} + \sqrt{\left(\frac{D}{A\omega^2}\right)^2 + 1}\right] \qquad (3.4\text{-}9)$$

The DF for the computer-and-relay combination is therefore given by

$$N(A,\omega) = \frac{4D}{\pi A} \exp\left[j \sin^{-1}\left(-\frac{D}{A\omega^2} + \sqrt{\left(\frac{D}{A\omega^2}\right)^2 + 1}\right)\right] \qquad (3.4\text{-}10)$$

which is amplitude- and frequency-dependent. Nevertheless, the frequency response of this time-optimal control system can now be obtained by application of the methods of Sec. 3.3.
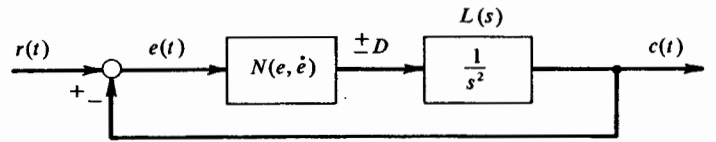
Figure 3.4-3   *Equivalent system for the frequency response analysis.*

Another avenue to the frequency response of this system was found by Sparacino (*op. cit.*), who observed that the open-loop transfer function of this system, $N(A,\omega)L(j\omega)$, could be written in terms of a *single* parameter, which we shall call F; namely,

$$N(A,\omega)L(j\omega) = \frac{4D}{\pi A\omega^2} \exp\left(j\left\{-\pi + \sin^{-1}\left[-\frac{D}{A\omega^2} + \sqrt{\left(\frac{D}{A\omega^2}\right)^2 + 1}\right]\right\}\right)$$

$$= \frac{4}{\pi F} \exp\left(j\left\{-\pi + \sin^{-1}\left[-\frac{1}{F} + \sqrt{\left(\frac{1}{F}\right)^2 + 1}\right]\right\}\right) \quad (3.4\text{-}11)$$

where

$$F = \frac{A\omega^2}{D} \quad (3.4\text{-}12)$$

This definition for F obviates the need for construction of a family of graphs in determination of the system frequency response, a quite desirable aspect. Such a manipulation is clearly not always possible.

We may now adopt the viewpoint that $N(A,\omega)L(j\omega)$, or, as it were, $NL(F)$, is a new nonlinearity N' in a closed-loop system wherein all linear
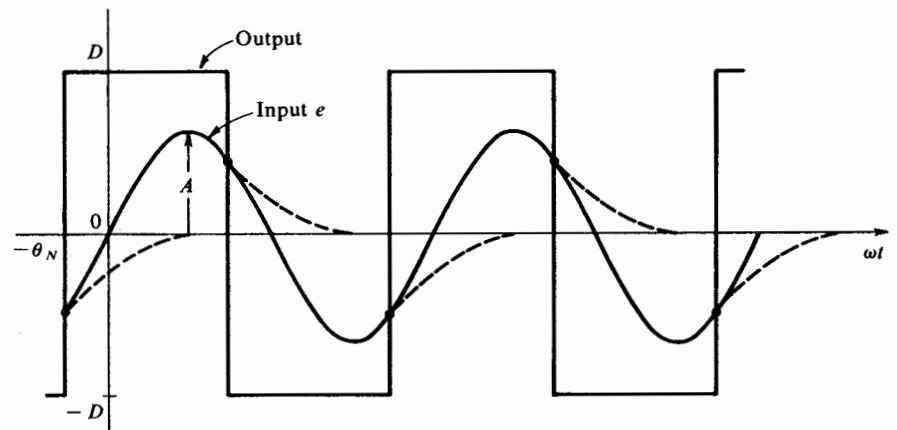

Figure 3.4-4   *Sinusoidal response of N for the time-optimal control of a pure inertia load.*

elements have unity amplitude transfer and zero phase shift. Thus, from Eqs. (3.3-5),

$$\rho_1 = \rho_2 = 1 \qquad \theta_1 = \theta_2 = 0$$

$$\rho_{N'} = \frac{4}{\pi F} \qquad \theta_{N'} = -\pi + \sin^{-1}\left[-\frac{1}{F} + \sqrt{\left(\frac{1}{F}\right)^2 + 1}\right] \tag{3.4-13}$$

Equation (3.3-6), governing the appropriate graphical construction, now becomes

$$\frac{A}{M_r} = \frac{1}{\sqrt{1 + 2\rho_{N'}\cos\theta_{N'} + (\rho_{N'})^2}} \tag{3.4-14}$$

where, as before, M_r is the magnitude of the system input sinusoid. Multiplying and dividing the left-hand member of Eq. (3.4-14) by ω^2/D,

$$\frac{A}{M_r} = \frac{F}{M_r \omega^2/D} \tag{3.4-15}$$

A plot of the left- and right-hand members of Eq. (3.4-14) yields solution points at observed intersections. The complicated right-hand member as a
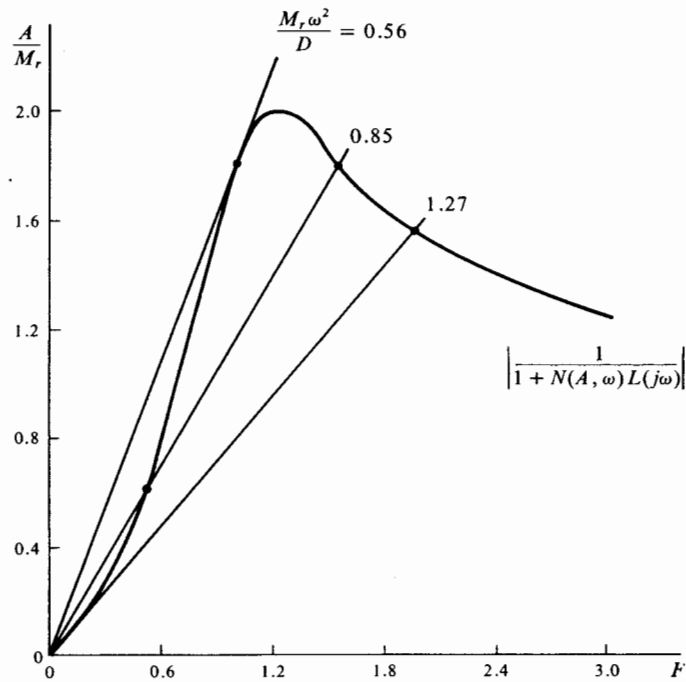

Figure 3.4-5 Graphical construction yielding A/M_r for a time-optimally-controlled pure inertia load. (Adapted from Sparacino, Ref. 96.)
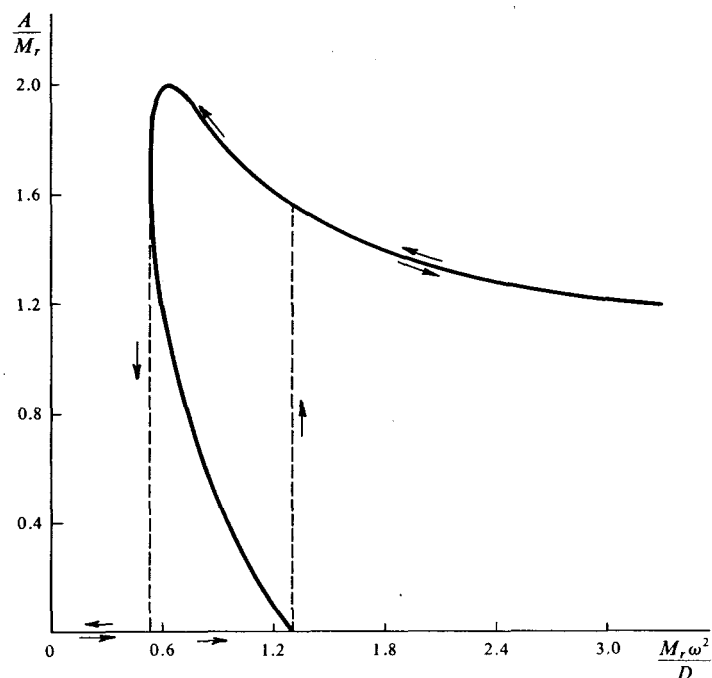
Figure 3.4-6 *Frequency-amplitude response of a time-optimally-controlled pure inertia load. (Adapted from Sparacino, Ref. 96.)*

function of F need be drawn but once, with the left-hand straight-line member drawn for a variety of values of $M_r\omega^2/D$, its reciprocal slope. Figure 3.4-5 illustrates these constructions. Note that $M_r\omega^2$ is the peak input acceleration and that, for the $1/s^2$ linear elements, D represents the peak output acceleration. Thus $M_r\omega^2/D$ is the ratio of peak commanded acceleration to peak output acceleration capability, a physically significant quantity. Changing the slope of the left-hand member of Eq. (3.4-14) therefore directly corresponds to sweeping out all the input dynamics–system dynamics possibilities. From another point of view, for fixed system dynamics the act of changing $M_r\omega^2/D$ directly corresponds to changing the command input. With either point of view the entire system harmonic response is immediately evident from the plot.

For values of $M_r\omega^2/D$ less than 0.56, one intersection occurs at $F = 0$. This implies that $A = 0$, or that the system follows the sinusoidal input with zero steady-state following error. For values of $M_r\omega^2/D$ between 0.56 and 1.27, three intersections occur. An abrupt transition from the previous state is indicated, with the system now displaying finite steady-state following error. For $M_r\omega^2/D > 1.27$, two intersections exist, one of which is nonzero. Figure 3.4-6 summarizes these results by displaying the error response vs. the

quantity $M_r \omega^2/D$.   Since it is both the frequency response (input amplitude constant) and amplitude response (input frequency constant), it has been so labeled.   Arrows indicate possible motion of the solution along the frequency-amplitude response locus.   The trivial solution for $M_r \omega^2/D > 1.27$ has been removed from expected possibilities, based on the physical reasoning that for $M_r \omega^2/D > 1$ a following error would have to exist.

Experimental results reported (Sparacino, *op. cit.*) confirm the presence of jump resonance as well as the entire frequency-amplitude response of Fig. 3.4-6.   The critical values of $M_r \omega^2/D$ so determined were approximately 0.60 and 1.15 as compared with 0.56 and 1.27 determined by the DF method. We may therefore conclude that the DF study predicted all the salient harmonic response behavioral patterns for the time-optimal controlled pure inertia load, with less than 10 percent error in determination of all critical parameter values.

3.5   LINEAR AND NONLINEAR LOOP COMPENSATION TECHNIQUES

The conventional compensation problem is to choose a network which, when inserted in a control loop, causes the overall loop (including fixed elements) to meet specifications.   One fixed element is usually the loop final element; that is, its output is the controlled variable $c(t)$.   Others, of course, are possible, any of which can be either linear or nonlinear.   An inherent loop nonlinearity is referred to as unintentional (or incidental); examples are output member backlash, power element saturation, and feedback element threshold.   Intentional nonlinearities include nonlinear compensation networks such as those discussed later on in this section.

Steady-state requirements for loop compensation commonly are (1) to rid an otherwise limit cycling system of its limit cycle (as in the case of a position servo with motor-load backlash), (2) to decrease the limit cycle amplitude to some acceptable value where its complete elimination is impossible (as in an on-off aircraft autopilot system), or (3) to permit control of the limit cycle amplitude and/or frequency (as in a vibration controller).   Compensation selection for steady-state requirements is readily accomplished via DF usage.

Transient response specifications such as percent overshoot and settling time are difficult to deal with in non-limit-cycling systems.   In totally linear systems it is well known that feedforward lead compensation tends to shorten the transient response time, whereas feedforward lag compensation increases transient response time.   Moreover, feedback lag compensation has an effect somewhat equivalent to feedforward lead, and vice versa for feedback lead compensation.   In non-limit-cycling nonlinear systems it is true that there

is often a certain qualitative agreement between the effect of linear compensation on the actual transient response and that described above for the totally linear case.   Because of the inapplicability of linear superposition to these nonlinear systems, however, such *equivalences* are weak, and should at most be used as a rough guide in design.   Chapters 4 and 8 are directed toward the study of transient phenomena, and further inferences regarding compensation for transient response can be drawn from material presented there.

Transient response characteristics of limit cycling systems are readily compensated, as we shall find in Chap. 6.   With the aid of a dual-input describing function, it will be shown there that the above-mentioned equivalences are no longer weak.   The reader is asked to temporarily accept this statement without proof.

Many theories are available to assist in the choice of loop compensation for totally linear systems.   Such is not the case for nonlinear systems, especially of third and higher order.   In this section we briefly examine several means of compensating these nonlinear systems, using DF theory as a guide to compensation selection.

LINEAR COMPENSATION

Here we presume the existence of a nonlinear fixed element, unintentional or otherwise, and inquire of the use of linear loop compensation.

The steady-state effects of both lead and lag networks on a linear system are conveniently displayed on the amplitude-phase plane, where linear compensation is easily carried out.   Let us assume that a system with a third-order transfer function also contains an odd memoryless nonlinearity, whose DF lies on the $-180°$ line in the amplitude-phase plane (see Fig. 3.5-1). The compensation network has the transfer function

$$L_1(s) = \frac{\alpha\tau s + 1}{\tau s + 1} \qquad \text{lead network for } \alpha > 1, \text{ lag network for } \alpha < 1$$

$$(3.5\text{-}1)$$

for which the insert in Fig. 3.5-1 illustrates the frequency locus in log magnitude vs. phase coordinates.   We recall that in the process of linear compensation of linear systems the lag network is used for its *attenuation* characteristic at high frequencies (relative to $1/\alpha\tau$), whereas the lead network is used for its *phase-lead* characteristic at low frequencies (relative to $1/\tau$).   These portions of the amplitude-phase loci of $L_1(j\omega)$ have been darkened in the insert.

The limit cycle resulting from the use of $L_1(s)$ as a *lead* network has a *higher frequency* and *smaller amplitude* than the limit cycle in the uncompensated system (A_0,ω_0).   The lead network has been chosen to supply maximum phase lead in the vicinity of ω_0.   When $L_1(s)$ is a *lag* network, a (slightly) *lower-frequency* and *smaller-amplitude* limit cycle is seen to result.   The lag
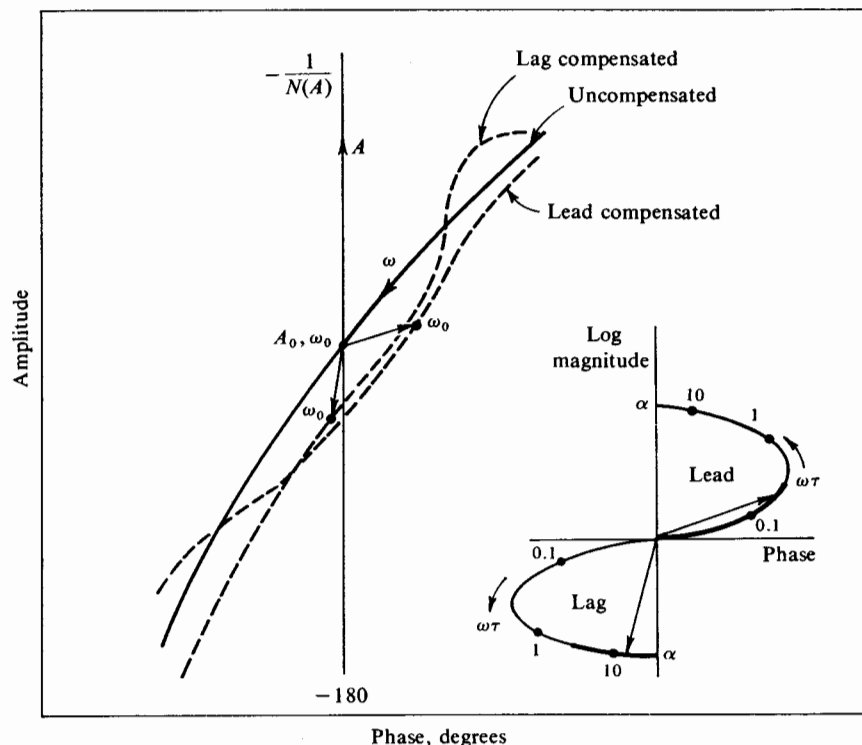
Figure 3.5-1 *Effect of linear lead and lag compensation networks viewed on the amplitude-phase plane.*

network chosen has its upper breakpoint well below ω_0, so that in the vicinity of ω_0, only its attenuation, and not its phase lag, predominates.

Quite commonly the system designer is constrained to place a compensation network at locations between the system output and the nonlinearity input.   In a control loop for a pulse-rebalanced accelerometer, for example, compensation *must* be placed in the feedback path since the feedforward path is physically inaccessible.   Referring to Fig. 3.5-1, if we assume $L_1(s)$ to be located between the system output and the nonlinearity input, it follows that the limit cycle amplitude *at the system output* is smaller when lead compensation is used than when lag compensation is used.

To see this, recall first that the limit cycle amplitude solved for graphically is that which appears at the nonlinearity *input*.   The amplitude at other stations around the loop must be found by working forward or back from this point.   Also, the lead network presents a *higher-than-unity gain* to the limit cycle, whereas the lag network presents a *lower-than-unity gain*.   Thus, for roughly equal limit cycle amplitudes at the nonlinearity input (as indicated

by Fig. 3.5-1), it must follow that the output limit cycle amplitude is smaller in a lead-compensated system.

The effect of a linear gain K is simply to raise or lower a frequency response locus on the amplitude-phase plane. This operation, which does not affect limit cycle frequency in the case of non-phase-shifting DFs, is shown in Fig. 3.5-2.

Between the use of linear lead, lag, and pure gain networks, limit cycle amplitude and frequency can theoretically be altered as desired to meet specifications. The prevention of limit cycles (where possible) without unnecessarily dropping the loop gain is usually a matter of supplying sufficient phase lead to avoid the crossing of the $-1/N(A,\omega)$ locus by the $L(j\omega)$ locus (see Fig. 3.5-3). Because of familiar noise and saturation considerations, however, a practical upper limit on α of the lead compensation network is about 10.

An additional linear compensation technique occasionally used to achieve a specified limit cycle amplitude at a particular point in a control loop is to
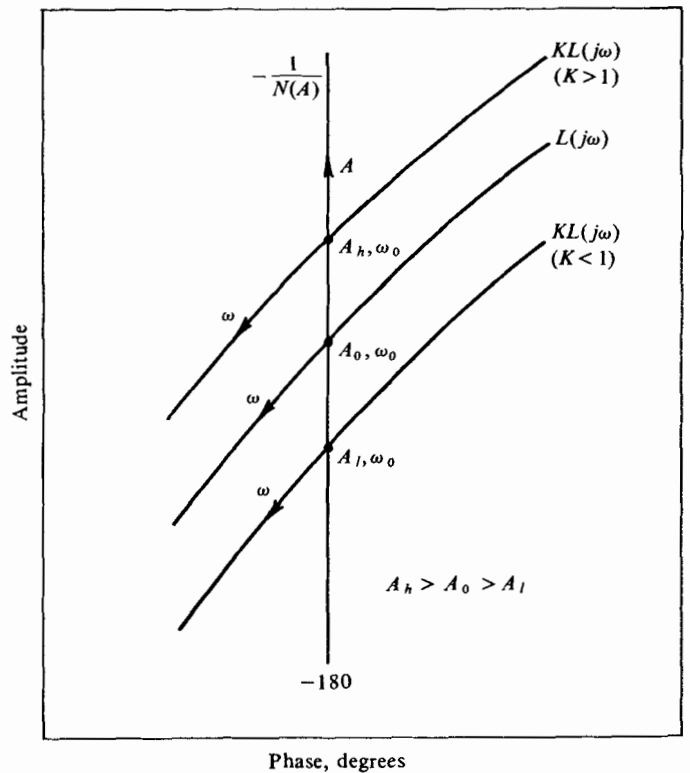

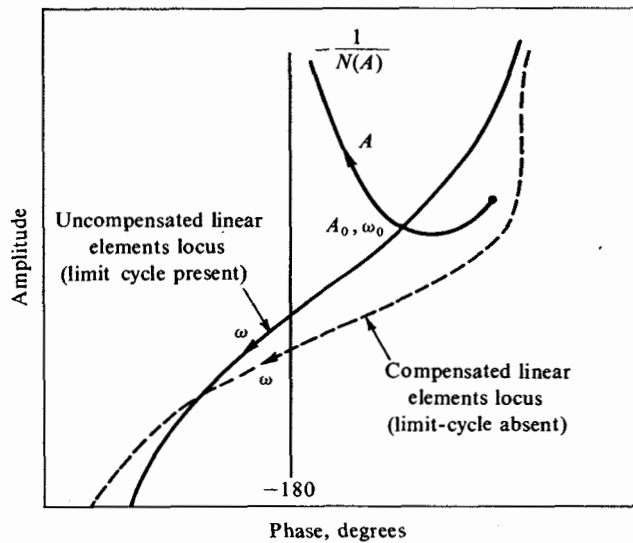Figure 3.5-2   Effect of linear gain.

Figure 3.5-3   Lead compensation to remove a limit cycle in a high-order system containing a relay with dead zone and hysteresis.

insert a lead (lag) network, followed elsewhere by the reciprocal lag (lead) network.   The net effect of this reciprocal network pair is to alter the limit cycle amplitude at all stations between the networks, but not its frequency. If one network is in the forward path and its reciprocal is in the feedback path, loop command response characteristics are clearly affected and must be accounted for.

NONLINEAR MODEL FEEDBACK

One of the interesting procedures available for synthesizing a nonlinear compensation network is based on what has been called complementary model feedback (Ref. 73).   In its basic form, nonlinear model feedback serves to linearize a nonlinear feedback system by providing a signal, which, when added to the nonlinearity output, renders its net output a linear function of its net input.   The addition of signals does not actually take place at the nonlinearity output, however, since that would require a nonlinear compensation network operating at high power levels in all cases in which the original nonlinearity is a power element.   The signal addition, as a result, conceptually takes place in the feedback path (Fig. 3.5-4b), although it can be implemented in the feedforward path (Fig. 3.5-4c)

Consider the case of *exact* model feedback, $N^* = N$, $L_2^* = L_2$.   The complementary nonlinearity $1 - N^*$ becomes $1 - N$.   In the case where N

(a)

(b)

(c)
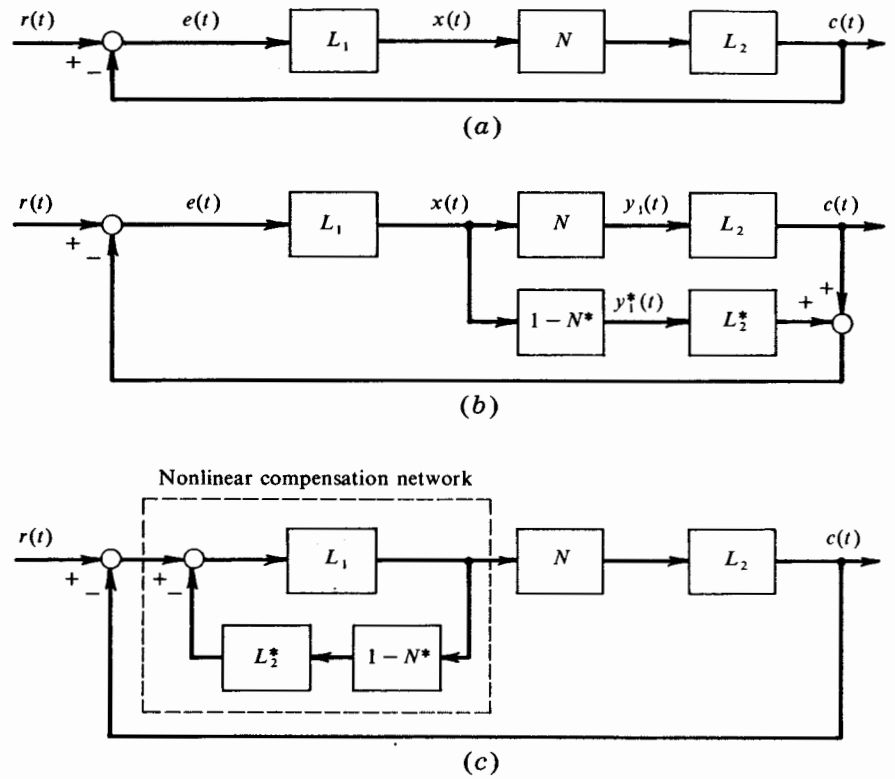
Figure 3.5-4 *Nonlinear model feedback. (a) Original system; (b) system with model feedback; (c) equivalent nonlinear compensator.*

is an abrupt saturation nonlinearity, for example, $1 - N$ is a piecewise-linear dead zone, and vice versa. It is clear that

$$Y_1(s) + Y_1^*(s) = X(s) \qquad (3.5\text{-}2)$$

which implies that the transfer function from $X(s)$ to $E(s)$ is independent of N, viz.,

$$\frac{E}{X}(s) = -\frac{Y_1(s)L_2(s) + Y_1^*(s)L_2^*(s)}{X(s)}$$

$$= -\frac{[Y_1(s) + Y_1^*(s)]L_2(s)}{X(s)}$$

$$= -L_2(s) \qquad (3.5\text{-}3)$$

Stability of this totally *linear* loop is therefore governed by the equation

$$1 + L_1(s)L_2(s) = 0 \qquad (3.5\text{-}4)$$

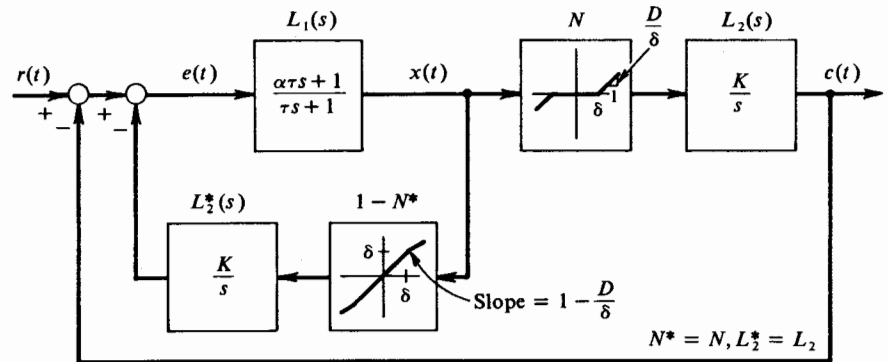*Figure 3.5-5* *System requiring the use of inexact model compensation.* (*Adapted from Mishkin and Braun, Ref. 72, p. 209.*)

independent of the loop nonlinearities. DF approximations have not, as yet, entered our discussion. The exact output in response to any input is obtained by solving for $x(t)$ *on a purely linear basis,*

$$x(t) = \mathscr{L}^{-1}[X(s)] = \mathscr{L}^{-1}\left[\frac{L_1(s)R(s)}{1 + L_1(s)L_2(s)}\right] \tag{3.5-5}$$

operating on the resulting time function by N, and passing the output of N through $L_2(s)$.

It is possible that either or both of N^* and L_2^* will be different from N and L_2, respectively. This can be due either to an *inaccuracy* in the model elements or to the use of an *intentionally* inexact model (Ref. 72). As an example of a situation where an inexact model should be intentionally employed, consider the system of Fig. 3.5-5, subject to a ramp input $r(t) = Mt$. As a consequence of the integration in $L_2(s)$ [hence in $L_2^*(s)$], the output signal $c(t)$ will remain *at rest* if the following relationship holds:

$$K\delta \geqslant M \tag{3.5-6}$$

for in this instance a *constant* value of x less than δ results, which never exceeds the threshold of N. The difficulty, of course, is due to the integration in $L_2^*(s)$. System response degradation is perhaps at an all-time high; the use of an inexact model is certainly indicated. A first choice for $L_2^*(s)$ might be an *approximate* integration, $L_2^*(s) = K/(\tau s + 1)$. Certainly the steady-state ramp response problem would be avoided. Note that this example points out the need for calculation of the system input-output response properties, independent of the fact that the loop transfer may be linear and possess suitable stability.

To investigate loop stability under any of the circumstances for which $N^* \neq N$ and/or $L_2^* \neq L_2$, we employ the DF representation for N and

$1 - N$, which we denote, as usual, by $N(A,\omega)$ and $1 - N(A,\omega)$, respectively. Representing the errors and/or inaccuracies in L_2^* and $1 - N^*$ by additive terms (denoted by Δ) as follows:

$$L_2^*(s) = L_2(s) + \Delta L_2(s)$$
$$1 - N^*(A,\omega) = 1 - N(A,\omega) - \Delta N(A,\omega)$$

(3.5-7)

it is easily demonstrated with reference to Fig. 3.5-4b that loop stability is now governed by the equation

$$1 + L_1(s)L_2(s)\left\{1 - \Delta N(A,\omega) + \frac{\Delta L_2}{L_2}(s)[1 - N(A,\omega) - \Delta N(A,\omega)]\right\} = 0$$

(3.5-8)

The long term in braces can be treated as some new amplitude- and frequency-dependent nonlinearity, with DF analysis proceeding as described in previous sections of this chapter.

Generally speaking, the design with exact models is unnecessary, perhaps difficult, and can lead to severely degraded system performance unless care is taken to avoid situations such as the one described above. The use of intentionally inexact models, however, represents an attractive area for non-linear compensation.

OTHER NONLINEAR NETWORKS

Linear minimum-phase networks are such that input-output gain and phase relationships are themselves related. Therefore choice of one constrains the form of the other. Linear non-minimum-phase networks display more phase lag than the corresponding minimum-phase networks (i.e., for the same attenuation characteristics), and thus they are of little use in system compensation. Nonlinear networks can be synthesized, however, where DF gain and phase relationships are chosen separately to suit the designer (Ref. 59). An example of one such circuit is the nonlinear Clegg integrator discussed in Sec. 2.4. It was shown to have a DF with an amplitude vs. frequency characteristic identical with that of a linear (minimum-phase) integrator and an associated DF phase of $-38°$ as opposed to $-90°$ for the linear integrator. This phase characteristic is certainly a desirable feature for loop compensation purposes, in which excessive phase lag can cause degraded or even unstable system behavior.

Systems with saturation in a tachometer feedback path may be unstable at large-feedback-signal levels. A useful compensation network would reduce the loop gain to large signals. Large-amplitude oscillations can occur in a conditionally stable system (where a gain reduction leads to instability). This behavior can be compensated by the use of a nonlinear network which provides phase lead at large-signal levels. Systems with

backlash all too often display low-amplitude limit cycles. A network providing phase lead at low-amplitude signals can be used for stabilization in these instances. Networks providing phase lead at low amplitudes are also useful in the stabilization of systems with spring-coupled coulomb friction (Ref. 60). Mechanizations for the three nonlinear networks indicated above are illustrated in Fig. 3.5-6. Note that the mechanizations indicated are not unique; many other useful forms can be developed.

The first of the networks illustrated is simply a gain-changing element drawn in a feedback mechanization. As such, the DF for this network can be derived directly as shown in Sec. 2.3. We shall present an approximate DF derivation, however, which will find real use when applied to the non-linear networks of Fig. 3.5-6b and c, for which the exact DF cannot be

$$x < \frac{\delta}{K} \quad \Rightarrow \quad \frac{Y}{X}(s) = K$$

$$x >> \frac{\delta}{K} \quad \Rightarrow \quad \frac{Y}{X}(s) \approx \frac{K}{1+K}$$

$$(a)$$

$$x < \frac{\delta}{K} \quad \Rightarrow \quad \frac{Y}{X}(s) = K$$

$$x >> \frac{\delta}{K} \quad \Rightarrow \quad \frac{Y}{X}(s) \approx \frac{K}{1+K}\left(\frac{\tau s+1}{\frac{\tau}{1+K}s+1}\right)$$

$$(b)$$

$$x < \frac{\delta}{K} \quad \Rightarrow \quad \frac{Y}{X}(s) = \frac{K}{1+K}\left(\frac{\tau s+1}{\frac{\tau}{1+K}s+1}\right)$$

$$x >> \frac{\delta}{K} \quad \Rightarrow \quad \frac{Y}{X}(s) \approx K$$
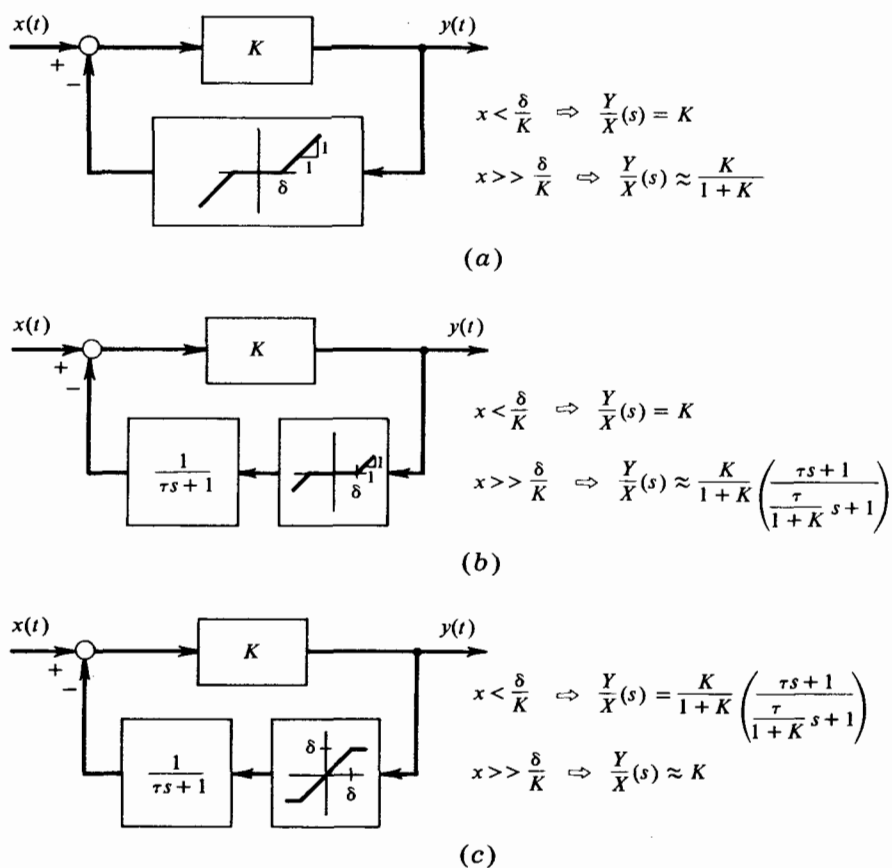
$$(c)$$

Figure 3.5-6 *Nonlinear networks providing (a) gain reduction at large-signal levels; (b) phase lead at large-signal levels; (c) phase lead at small-signal levels.*

found in any convenient analytic way.   Let $N^\circ(A)$ represent the DF for the feedback dead-zone element in Fig. 3.5-6a.   Now assume a sinusoidal *output* (we have already used this artifice in the study of implicit dynamic non-linearities in Sec. 2.4)

$$y = A \sin \omega t \qquad (3.5-9)$$

Since for this network there is no internal phase shift, a first-harmonic loop balance (ignoring fed-back harmonics) is achieved when

$$[A_{in} \sin \omega t - AN^\circ(A) \sin \omega t]K = A \sin \omega t \qquad (3.5-10)$$

from which the frequency-invariant input-output DF is found to be

$$N(A) = \frac{A}{A_{in}}$$
$$= \frac{1}{N^\circ(A) + 1/K} \qquad (3.5-11)$$

Several observations are in order.   First, the DF is a function of A (at the output) as opposed to A_{in}.   This presents no particular difficulty in its use, of course; one need only recognize that a given magnitude of the DF corresponds to a particular value of A.   Second, the formulation, although approximate, does yield exact results in the limiting cases $A_{out} \to 0$ [where $N(A_{out}) = K$)] and $A_{out} \to \infty$ [where $N(A_{out}) = K/(1 + K)$], with small errors elsewhere.   For example, in the case where $K = 1$ and $\delta = 1$, the maximum error in the approximate DF referred to the exact DF is less than 2 percent.   Third, and most important, in application of this DF the filter hypothesis must be reversed.   That is, for perfect DF results we *require* certain harmonics at x (the residual is now associated with the input), which must be generated by the passage of a pure sinusoid through a linear system (the loop linear elements).   Obviously, this is a physically unrealizable requirement.   Hence *perfect* results cannot be achieved using this DF model except at signal levels causing only linear operation.   Nonetheless, *useful* results can be achieved.   In this connection, Smith (Ref. 95, p. 461) points out that reversal of a cause-effect relationship is sometimes physically indicated.   For example, in the analysis of magnetic amplifiers with parallel ac windings or with a zero-impedance bias source, the flux is essentially sinusoidal.   Therefore the DF should be formulated in terms of flux input and magnetomotive force output, although the physical phenomenon is usually described as magnetomotive force for cause and flux for effect.

Perhaps the best means of determining $N(A,\omega)$ in the case of frequency-variant nonlinear networks is actual laboratory testing of a piece of hardware or a computer simulation of the network.   Alternatively, an approximate means of calculation such as demonstrated in this section can be employed.

Following the approximate DF calculation procedure, it is easily shown that the approximate DF for the network of Fig. 3.5-6b is given by

$$N(A,\omega) = \frac{1}{\sqrt{\left(\frac{1}{K}\right)^2 + \frac{2N^\circ(A)}{K(1 + \tau^2\omega^2)} + N^\circ(A)^2}}\ e^{j\,\tan^{-1}\left[\frac{KN^\circ(A)\tau\omega}{1+\tau^2\omega^2+KN^\circ(A)}\right]} \qquad (3.5\text{-}12)$$

Similarly, the approximate DF for the network of Fig. 3.5-6c is identical with the expression above, with the replacement of $N^\circ(A)$ for dead zone by $N^\circ(A)$ for saturation. Note the *phase lead* of each of these DFs, as required at either large- or small-signal levels.

3.6 TREATMENT OF MULTIPLE NONLINEARITIES

It is quite possible that nonlinearities will be present at more than one station around a control loop. In a typical active satellite attitude control system, for example, both the stabilizing (on-off) gas jets and the error transducer (perhaps a sun sensor) are quite decidedly nonlinear. In a synchro-controlled heavy-gun positioning system, both synchro error transducer and (saturating) hydraulic motor are again nonlinear, although in this case both nonlinearities are apparent only at large-signal levels. In fact, the particular combination of nonlinear error transducer and nonlinear power element are common occurrences in control-system applications. Thus it is fitting that we examine the possibility of using DF techniques in the study of multiple nonlinearity systems.

GENERAL MULTIPLE NONLINEARITY SYSTEMS

By a general two-nonlinearity system is meant one such as is illustrated in Fig. 3.6-1, where $N_1(x_1,\dot{x}_1)$ and $N_2(x_2,\dot{x}_2)$, as well as $L_1(s)$ and $L_2(s)$, are different. Let us replace the two nonlinearities by their respective DFs, $N_1(A_1,\omega)$ and $N_2(A_2,\omega)$. The frequency response of this system can be determined by solution of the equation

$$\frac{X_1}{R}(j\omega,A) = \frac{1}{1 + N_1(A_1,\omega)L_1(j\omega)N_2(A_2,\omega)L_2(j\omega)} \qquad (3.6\text{-}1)$$

provided that the use of the DF for linearization is permissible. Clearly, one requirement for this to be true is that $L_1(j\omega)$ and $L_2(j\omega)$ are *each* low-pass
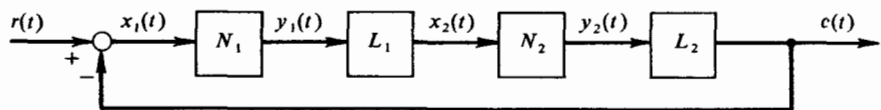

Figure 3.6-1 General two-nonlinearity system.

filters, so that the periodic inputs to both of the nonlinearities are nearly sinusoidal.

Equation (3.6-1) can be solved in a manner very similar to the first solution discussed in the one nonlinearity case, Sec. 3.3. By *choosing* values of A_1 and ω, $N_1(A_1,\omega)$ is determined and A_2 is, as a consequence, also determined.

$$A_2 = A_1 \, |N_1(A_1,\omega)| \, |L_1(j\omega)| \tag{3.6-2}$$

Thus the corresponding value of $N_2(A_2,\omega)$ is established, and Eq. (3.6-1) is soluble for X_1/R. This approach is in essence the same as considering the quantity $N_1(A_1,\omega)L_1(j\omega)N_2(A_2,\omega)$ equivalent to a new nonlinearity, $N(A_1,\omega)$, and proceeding as in earlier discussions of the single nonlinearity case.

Limit cycle study for this system corresponds to the determination of nontrivial solutions to the set of four relationships

$$
\begin{bmatrix}
N_1 & -1 & 0 & 0 \\
0 & L_1 & -1 & 0 \\
0 & 0 & N_2 & -1 \\
1 & 0 & 0 & L_2
\end{bmatrix}
\begin{bmatrix}
X_1 \\
Y_1 \\
X_2 \\
Y_2
\end{bmatrix}
= 0
$$

or equivalently,

$$
\begin{vmatrix}
N_1 & -1 & 0 & 0 \\
0 & L_1 & -1 & 0 \\
0 & 0 & N_2 & -1 \\
1 & 0 & 0 & L_2
\end{vmatrix}
= 0 \tag{3.6-3}
$$

which reduces to the single equation

$$1 + N_1(A_1,\omega)L_1(j\omega)N_2(A_2,\omega)L_2(j\omega) = 0 \tag{3.6-4}$$

Of course, we come to the same conclusion by setting the denominator of Eq. (3.6-1) to zero, corresponding to writing the characteristic equation for the purely linear case. Equation (3.6-4) can be solved by any of the techniques for limit cycle determination discussed in Sec. 3.1 by treating the quantity $N_1(A_1,\omega)L_1(j\omega)N_2(A_2,\omega)$ as an equivalent single nonlinearity $N(A_1,\omega)$. When $L_1(j\omega)$ is a low-pass function, this simplified approach will be in small error. On the other hand, should $L_1(j\omega)$ turn out to be a lead-lag or similar non-low-pass compensation network, the simplified approach could lead to large error.

A more accurate solution to both the frequency response and limit cycle problems can be had by computing the DF for the complete chain $N_1(x_1,\dot{x}_1)L_1(j\omega)N_2(x_2,\dot{x}_2)$. This is accomplished by determining the actual first harmonic in y_2 when x_1 is a pure sinusoid. It is, of course, a more difficult approach. Regarding this approach, observe that even if N_1 and

N_2 are only amplitude-dependent, N will be both amplitude- and frequency-dependent because of the presumed frequency dependence of L_1. Thus, in general, the treatment of multiple nonlinearity systems is bound to be somewhat more laborious than the treatment of single nonlinearity systems.

Example 3.6-1 Calculate the limit cycle frequency for the two-nonlinearity system of Fig. 3.6-2.

As N_2 is multiple-valued, it is convenient to seek a solution of Eq. (3.6-4) in the form

$$N_1(A_1)L_1(j\omega)L_2(j\omega) = - \frac{1}{N_2(A_2)}$$

The phase-shifting elements in this equation are $L_1(j\omega)L_2(j\omega)$ and $N_2(A_2)$. Thus, given values of ω must imply specific values of A_2 such that an appropriate loop phase shift occurs. Noting this phase requirement, we plot $L_1(j\omega)L_2(j\omega)$ and $-1/N_2(A_2)$ on amplitude-phase coordinates, as illustrated in Fig. 3.6-3. Any vertical line on this plot intercepts appropriate *pairs* of values (A_2,ω).

A_1 is now expressed as

$$A_1 = A_2 |N_2(A_2)| |L_2(j\omega)|$$

and is calculated for several pairs of the values (A_2,ω). This allows the calculation of several values of $|N_1(A_1)|$. Adjoining the resulting magnitudes (expressed in decibels, since this allows for graphical addition) to $L_1(j\omega)L_2(j\omega)$ at corresponding frequencies gives the *constructed locus*, whose intersections with $-1/N_2(A_2)$ satisfy both steady-state oscillation amplitude and phase-shift requirements, and hence yield the limit cycle solutions of interest. Figure 3.6-3 details the determination of one point on the constructed locus corresponding to the values $A_2 = 10$, $\omega = 3.9$.

By this process we observe as the limit cycle solution,

$$A_2 = 6.2 \qquad \omega = 3.55$$

This compares extremely well with the analog computer solution

$$A_2(\text{exper}) = 6.2 \qquad \omega(\text{exper}) = 3.49$$

particularly in view of the simplifying approximations employed.

LIMIT CYCLES IN SYMMETRIC MULTIPLE NONLINEARITY SYSTEMS

This section presents a method for transformation of a symmetric multiple nonlinearity system to a *single* nonlinearity system for the study of single-frequency limit cycle behavior. It is conjectured that there need be no
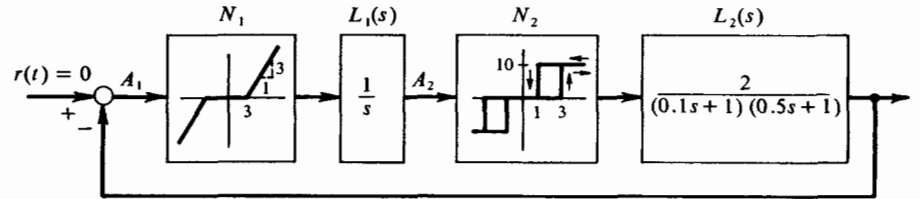

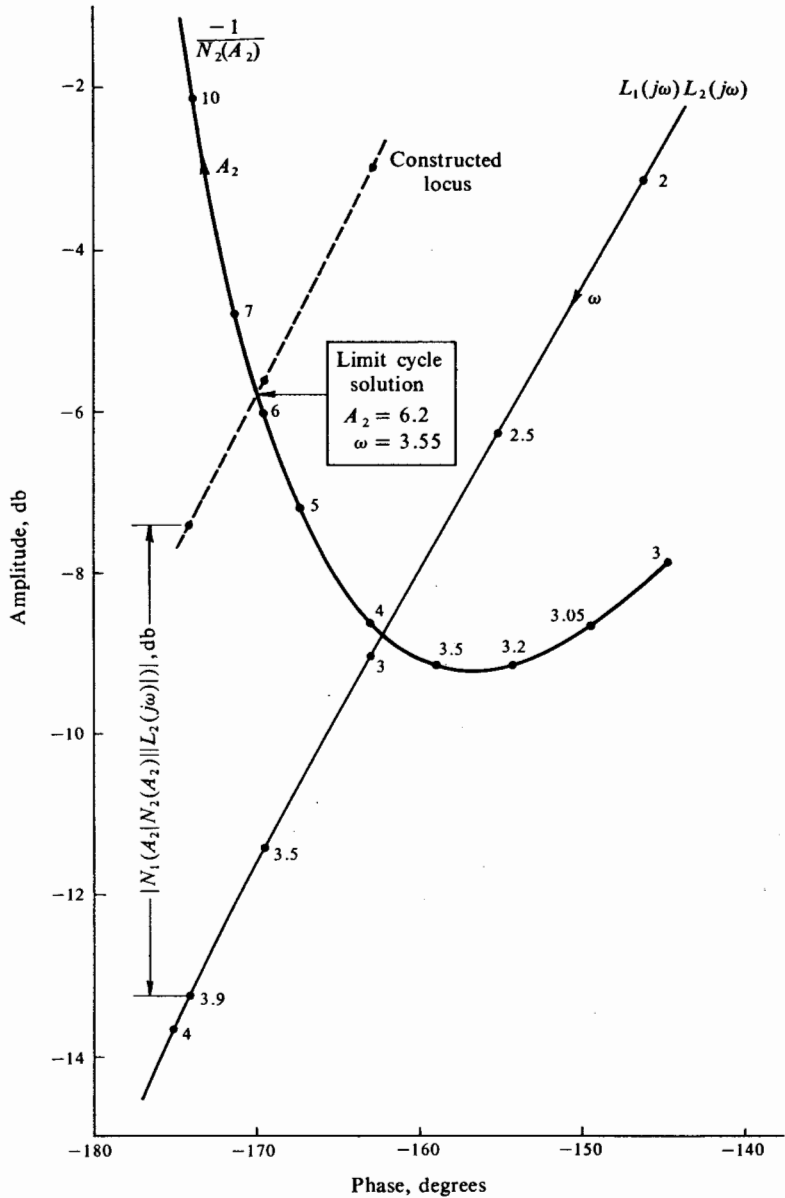*Figure 3.6-2* *Example two-nonlinearity system.*

Figure 3.6-3   *Graphical limit cycle solution in example two-nonlinearity system.*

No input → ⊕ → [N] → [L] — $x_1(t)$ → [N] → [L] → $x_2(t)$

(a)

○ → [N] → [L] — $x_1(t)$ → [$e^{-j\frac{\pi}{2}}$] → $x_2(t)$
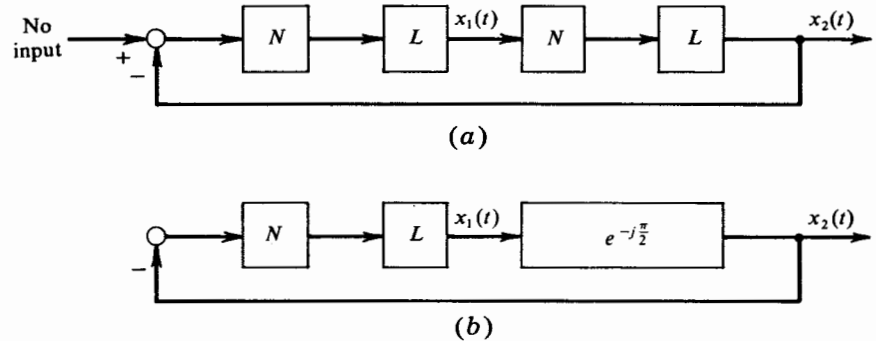
(b)

Figure 3.6-4 (a) *Symmetric two-nonlinearity single-loop system.* (b) *Single-nonlinearity equivalent for limit cycle study.*

restriction on the form of the linear elements, but that the nonlinearity is required to be odd. The resulting transformation, which apparently can be used for multiple- as well as single-loop systems, is exact. We first consider single-loop systems.

The essence of the matter is simply to observe that in a single-loop symmetric multiple nonlinearity system, any limit cycle which propagates must preserve its waveform at certain stations around the loop. Hence nonlinear operations on the limit cycle have a net effect *identical* with linear time delay (Ref. 27).

Consider the closed loop of Fig. 3.6-4. There is no input; we wish to study unforced oscillations. As the illustration implies, we have assumed a symmetric system in the sense that the two nonlinearities and the two linear elements are identical, respectively. It is plausible to argue that if the loop

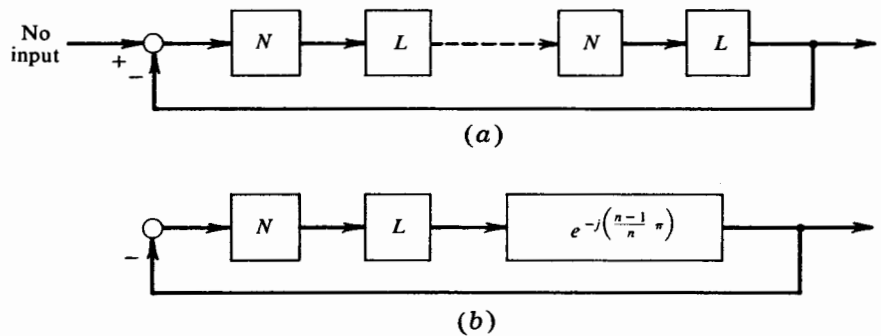No input → ⊕ → [N] → [L] — — — → [N] → [L] →

(a)

○ → [N] → [L] → [$e^{-j\left(\frac{n-1}{n}\pi\right)}$] →

(b)

Figure 3.6-5 (a) *Symmetric n-nonlinearity single-loop system.* (b) *Single-nonlinearity equivalent for limit cycle study.*

does support a limit cycle, then, except for phase, the waveforms $x_1(t)$ and $x_2(t)$ must also be identical.   In order to sustain a limit cycle we must have

$$x_1(t) = x_2\left(t + \frac{T_0}{4}\right) \tag{3.6-5}$$

where T_0 is the limit cycle period.   This follows from the requirement for a loop phase shift of $-360°$ in order to sustain an oscillation and the fact that the summing junction provides $-180°$ of phase.   Thus each of the pairs NL effectively provide a phase shift of one-quarter period, or $-90°$.   Hence it is argued that an equivalent representation for the system of Fig. 3.6-4a is as shown in Fig. 3.6-4b.   Further, it is argued that this representation must be exact for limit cycles of the assumed form.

By a continuation of the same argument, we conclude the results depicted in Fig. 3.6-5 for a symmetric n-nonlinearity single-loop system.

Example 3.6-2   Determine the functions ω_0 versus $\zeta(\omega_n = 1)$ and ω_0 versus $\omega_n(\zeta = 0.5)$ for a three-nonlinearity single-loop system with identical linear elements given by

$$L(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

and identical ideal-relay nonlinearities.

In the single nonlinearity equivalent for the three-nonlinearity case, the loop linear element is $L(s) \exp[-j(2\pi/3)]$.   Hence $L(j\omega_0)$ must supply $60°$ of phase lag, as follows:

$$\underline{/L(j\omega_0)} = -\tan^{-1}\frac{2\zeta\omega_n\omega_0}{\omega_n^2 - \omega_0^2} = -\frac{\pi}{3}$$

Solving for ω_0 yields

$$\omega_0 = \frac{\zeta\omega_n}{\sqrt{3}}\left[\sqrt{1 + \left(\frac{3}{\zeta^2}\right)} - 1\right]$$

Theoretical and experimental results are plotted in Fig. 3.6-6.

As can be seen, the theoretical results are quite good.   By observing simulated outputs of the linear elements, the theoretical phase shift of $-60°$ per block (NL constituting a block) is verified.   The slight discrepancy between theoretical and experimental results for high ζ is a consequence only of the approximations in describing function theory (i.e., waveforms in this region are highly nonsinusoidal).   Exact analyses of the equivalent single-nonlinearity systems must lead to the exact limit cycle predictions.   During simulation it was noted that all results were initial-condition-independent, as expected.

It is possible to extend these results to the cases of coupled multiple-loop systems.   Linearly coupled nonlinear loops and nonlinearly coupled linear loops can both be studied, although only the latter case leads to unique results.   Before discussing the possibility of nonuniqueness, let us first obtain the equivalent single-nonlinearity system.

Consider the system of Fig. 3.6-7a, where L_1 and L_2 are linear networks and N is a nonlinear operator.   The symmetry in this system is apparent.
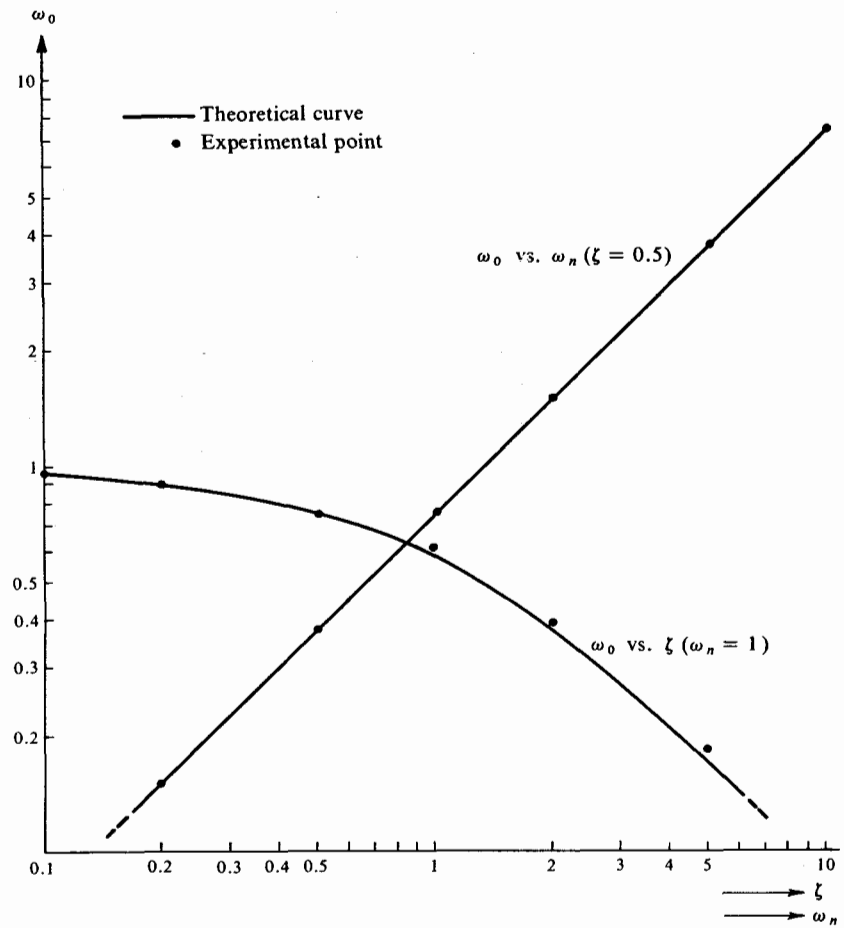
Figure 3.6-6 Experimental results for a symmetric three-nonlinearity single-loop system. $\omega_0 =$ limit cycle frequency.

Following the argument presented earlier, we should assert that $x_1(t)$ and $x_2(t)$ are identical in waveshape, but possibly of different phase. Since the phase shift from x_1 to x_2 and back to x_1 must be $-360°$ or a multiple thereof, and since the phase shift from x_1 to x_2 is identical with that from x_2 to x_1, we should argue that in order for the coupled system to oscillate at one frequency with a repeated waveshape, the phase shift from x_1 to x_2 must be a multiple of $-180°$. Hence we obtain the equivalent system of Fig. 3.6-7b. It is of the single-loop variety, with an equivalent linear element $L_1(s) \pm L_2(s)$.

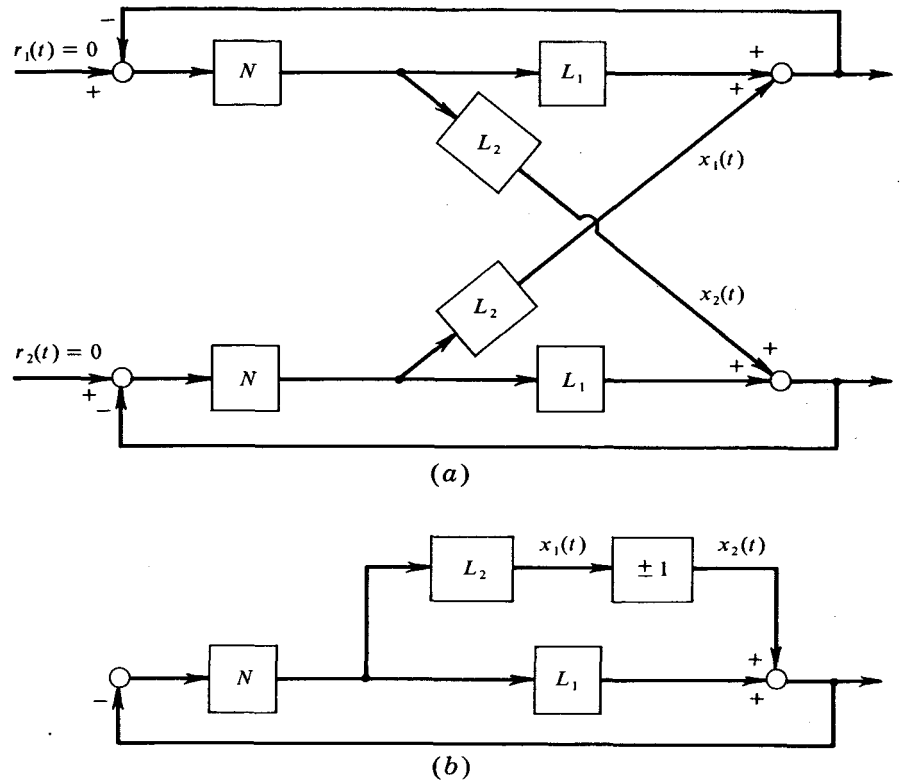Other cases can be treated in a similar manner. Nonlinearly coupled

Figure 3.6-7   (a) *Symmetric two-nonlinearity multiple-loop system.*   (b) *Single-nonlinearity equivalent for limit cycle study.*

linear loops, for example, degenerate to the single-loop case presented in the first part of this discussion, provided that one of the coupling elements contains a sign-changing term.   In this situation any limit cycles which occur *must* couple both linear loops, and the results so obtained are unique.

As an example of a possibility for which this model is incomplete, consider the case where N and $L_1(s)$ are of such forms that each of the two loops present when $L_2(s) = 0$ can sustain two or more stable limit cycles. Now let us take the situation where both loops are supporting different limit cycles and $L_2(s)$ is small.   The equivalent representation of Fig. 3.6-7b fails to account uniquely for this situation, indicating only that two stable limit cycles can indeed occur for sufficiently small values of $L_2(s)$.   The equivalent representation does allow study of limit cycle behavior, but it excludes some of the possible limit cycle cases.

3.7 ACCURACY OF THE DF APPROXIMATION[1]

As is so often the case in work with various types of describing functions, accuracy estimators are not part of the theoretical bundle from which the describing function itself is formulated. Additional means must be sought out, with which to provide accuracy estimates. These schemes must inevitably take cognizance of the *residual*, that part of the nonlinearity output ignored in the actual describing function development. In the case of the sinusoidal-input describing function, the residual consists of the entire sinusoidally forced nonlinearity output, minus the fundamental component. These harmonics, along with the linear elements, determine whether, and to what degree, DF solution of a problem will be successful.

In a number of text examples thus far, we have seen that results obtained through use of the DF approximation have been quite successful from a qualitative point of view, and that quantitative results have as well been generally excellent. It is also possible to demonstrate that both qualitative and quantitative conclusions can be in gross error, as shall be done later in this section. By examination of a few selected examples, we shall see that satisfaction of the *filter hypothesis* requirement is indeed the keystone of DF success. Moreover, we shall be in a position to estimate the accuracy of DF

---

[1] The text discussion throughout this chapter is based on the *filter hypothesis*, namely, that the loop linear elements attenuate nonlinearity output harmonics to the point that the nonlinearity input waveform is very nearly sinusoidal. Aizerman (Ref. 1) points out that for nonlinearities characterized by $y = mx + \mu\varphi(x)$, where μ is a small parameter, if for $\mu = 0$ the resulting closed-loop system has a complex pair of very lightly damped poles, the nonlinearity input will also be very nearly sinusoidal. He calls these the conditions of the *autoresonance hypothesis*.

In the case of autoresonance we can postulate beforehand that the limit cycle frequency is very nearly equal to the natural frequency of the lightly damped pole pair. Call this frequency ω^*. Then a small correction to ω^*, called $\delta\omega$, is found from the expansion of

$$L(j\omega^* + j\delta\omega) = -\frac{1}{N(A)}$$

in the form

$$L(j\omega^*) + \frac{dL(j\omega)}{d\omega}\bigg|_{\omega=\omega^*}\delta\omega = -\frac{1}{N(A)}$$

Graphical solution of this equation is readily implemented.

Aizerman (*op. cit.*) argues that when $N(A)$ is real, $\delta\omega = 0$, and both the *autoresonance hypothesis* and the *filter hypothesis* lead to identical results. However, when $N(A)$ is a complex function of A, the employment of the *autoresonance hypothesis* in the study of systems which satisfy the *filter hypothesis* can lead to considerable error. On the other hand, use of the *filter hypothesis* in situations where the *autoresonance hypothesis* is satisfied does not lead to contradictory results, and in fact gives a more accurate solution. For these reasons and the additional fact that in problems of automatic control the *autoresonance hypothesis* is rarely satisfied, we shall pursue this topic no further.

results a priori, an essential feature of DF theory, based upon casual inspection of system linear elements and nonlinearities.

An independent discussion of the filter hypothesis is presented in Appendix G.

IMPORTANCE OF HARMONICS

This topic is profitably pursued first by example, and later by discussion. We shall examine five different systems in some detail, emphasizing the role played by harmonics in DF success or failure.

The first three examples are studies of a conservative second-order system with various nonlinearities: piecewise-linear *preload*, *harmonic*, and *odd polynomial* characteristics, in that order. The first example could arise physically from an imperfect electronic amplifier simulation (finite output impedance) of an ideal relay controlling a motor of negligible time constant with a pure inertia load. The second example is a study of large-amplitude oscillations of an ideal pendulum. The last of the first three examples could have its physical origin in the behavior of a mass on a nonlinear polynomial spring. Exact solutions are derived, along with DF solutions for these examples, in order that percent error in DF utilization can be established. The fourth example is a brief study of a pulse-excited damped clock pendulum, with the exact solution also given. The last example concerns a temperature control loop and the apparent reasons for which DF analysis fails to predict behavior of this system appropriately.

Example 3.7-1   Determine the free-oscillation relationships which exist in the conservative second-order nonlinear system of Fig. 3.7-1. $N(x)$ is the *preload* nonlinearity.

The differential equation governing the behavior of this system is

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

By defining a new quantity $z = \dot{x}$ and writing $\ddot{x}$ in the form $z\, dz/dx$, we can separate variables to obtain

$$z\, dz = -y(x)\, dx$$

Integrating this result from A_0, the peak value of x at which z is zero, to the literal value x, we get

$$\frac{z^2}{2} = -\int_{A_0}^{x} y(x)\, dx \qquad \text{or} \qquad z = \left[2\int_{x}^{A_0} y(x)\, dx \right]^{\frac{1}{2}}$$

If the nonlinearity is odd, the free-oscillation waveform will be symmetric with respect to a quarter cycle.   Using the above solution for z, we can establish T_0, the oscillation period, as four times the interval during which x grows from 0 to A_0.

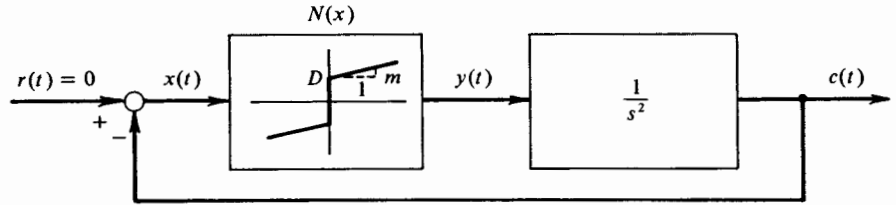$$T_{0(\text{exact})} = 4\int_{0}^{A_0} \left[2\int_{x}^{A_0} y(x)\, dx \right]^{-\frac{1}{2}} dx$$

Figure 3.7-1   *Nonlinear conservative second-order system.*

The free-oscillation frequency is therefore determined exactly as

$$\omega_{0(exact)} = \frac{2\pi}{T_{0(exact)}} = \frac{\pi}{\sqrt{2}} \left\{ \int_0^{A_0} \left[\int_x^{A_0} y(x)\, dx \right]^{-\frac{1}{2}} dx \right\}^{-1} \tag{3.7-1}$$

We can now specialize this result to the case of a preload nonlinearity.   Proceeding as required,

$$\int_x^{A_0} y(x)\, dx = \int_x^{A_0} \left(mx + D \frac{x}{|x|} \right) dx$$

$$= \frac{m}{2} A_0{}^2 + D\,|A_0| - \frac{m}{2} x^2 - D\,|x|$$

which yields, as the solution for ω_0

$$\omega_{0(exact)} = \frac{\pi}{\sqrt{2}} \left[\int_0^{A_0} \frac{dx}{\sqrt{(m/2)A_0{}^2 + DA_0 - (m/2)x^2 - Dx}} \right]^{-1}$$

$$= \frac{\pi}{\sqrt{2}} \left[\frac{\sqrt{2}}{\sqrt{m}} \sin^{-1} \left(\frac{mx + D}{mA_0 + D} \right) \Big|_0^{A_0} \right]^{-1}$$

$$= \frac{\sqrt{m}}{1 - \dfrac{2}{\pi} \sin^{-1} \left(\dfrac{1}{1 + (m/D)A_0} \right)} \tag{3.7-2a}$$

Observe that this system does not *limit-cycle;* in contrast, it supports a *conservative free oscillation*, the difference between the two being that whereas the former corresponds to a single *discrete* equilibrium state of the system over some finite range of initial conditions, the latter corresponds to a continuous *spectrum* of equilibrium states dependent directly on the initial conditions.

For comparison let us obtain the frequency of oscillation by DF usage.   In DF application the differential equation of motion of the system is written in the linear form

$$\ddot{x} + N(A)x = 0$$

from which it is immediately evident that the frequency of oscillation is

$$\omega_{0(DF)} = \sqrt{N(A_0)}$$

$$= \sqrt{\frac{4D}{\pi A_0} + m}$$

$$= \sqrt{m} \sqrt{1 + \frac{4}{\pi} \frac{D}{mA_0}} \tag{3.7-2b}$$

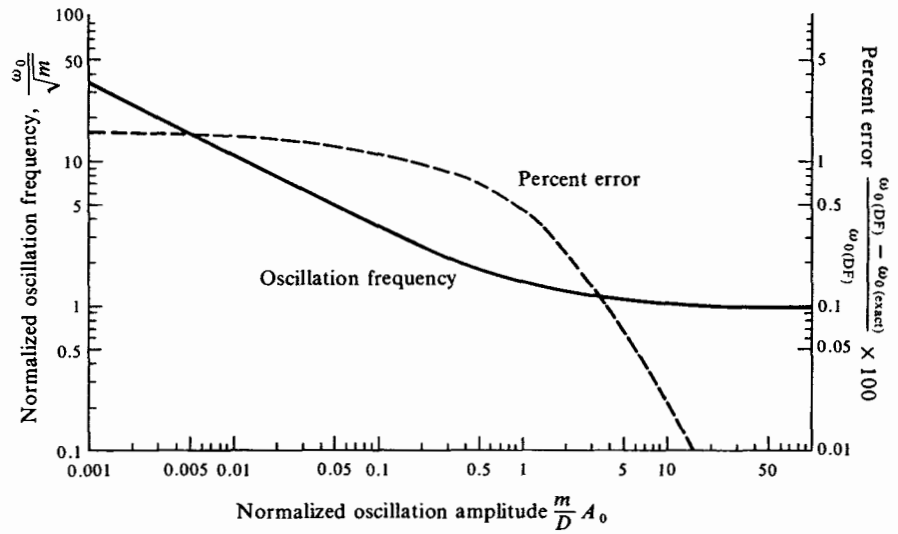where $N(A)$ is as calculated in Sec. 2.3.

Figure 3.7-2 *Conservative-free-oscillation frequency vs. amplitude in a doubly integrating second-order system with a preload nonlinearity.*

Figure 3.7-2 is a plot of $\omega_0/\sqrt{m}$ for both the DF and exact oscillation frequency calculations. The maximum error in DF approximation over the entire range of mA_0/D is 1.6 percent, occurring in the limit as $mA_0/D \to 0$. In this region the preload nonlinearity is indistinguishable from an ideal relay. As $mA_0/D \to \infty$, the purely linear case is approached, and the error goes to zero. The free-oscillation frequency curves are indistinguishable.

Example 3.7-2 Determine the oscillation frequency of a simple pendulum of length l in a constant gravity field g.

The equation with which we are concerned is well known, namely,

$$\ddot{x} + \frac{g}{l}\sin x = 0$$

With the aid of the identity

$$\cos x = 1 - 2\sin^2\frac{x}{2}$$

and the definition for φ,

$$\sin \varphi = \csc\frac{A_0}{2}\sin\frac{x}{2} \qquad \begin{cases} x = A_0 \Rightarrow \varphi = \dfrac{\pi}{2} \\ x = 0 \Rightarrow \varphi = 0 \end{cases}$$

the exact solution can be obtained from Eq. (3.7-1), in the form

$$\omega_{0(\text{exact})} = \frac{\pi}{2}\sqrt{\frac{g}{l}}\left[K\left(\sin\frac{A_0}{2}\right)\right]^{-1} \tag{3.7-3a}$$

where $K(\text{arg})$ is the complete elliptic integral of the first kind.   For $k < 1$,

$$K(k) = \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1 - k^2 \sin^2 \varphi}}$$

Another form of this integral appears in Eq. (2.3-29).

For comparison, the DF solution is obtained directly by use of Eq. (2.3-23).

$$\omega_{0(\text{DF})} = \sqrt{\frac{g}{l}} \sqrt{N(A_0)}$$

$$= \sqrt{\frac{g}{l}} \sqrt{\frac{2J_1(A_0)}{A_0}} \tag{3.7-3b}$$

where $J_1(A)$ is the Bessel function of order 1 and argument A.

The normalized frequency of conservative free oscillations obtained by each of these methods is plotted versus A_0 in Fig. 3.7-3.   The percent error in approximation is also shown.   From this curve we see that for oscillation amplitudes less than 75° the error in approximation is 0.1 percent.   By way of contrast, the error at 75° obtained by linearizing the sine function for small angles ($\sin x \sim x$) is 10 percent.   At an amplitude of 130°, the error in DF approximation is 2 percent, and the linear approximation is in error by 32 percent.

For completeness we mention that the DF-predicted oscillation amplitudes for which the linearly predicted oscillation frequencies are in error as indicated above are themselves in error by 1 percent (at an exact amplitude of 75°) and 4 percent (at an exact amplitude of 130°).

Example 3.7-3   Compare the DF and exact free-oscillation frequency solutions for a conservative second-order system with a one-term odd polynomial nonlinearity.

The odd nonlinearity is defined by

$$y = \begin{cases} x^n & n \text{ odd} \\ x^{n-1} |x| & n \text{ even} \end{cases}$$

Proceeding as in the previous examples, the exact solution is found in the form

$$\omega_{0(\text{exact})} = \frac{\pi}{\sqrt{2}} \left(\int_0^{A_0} \frac{dx}{\sqrt{[1/(n + 1)](A_0^{n+1} - x^{n+1})}} \right)^{-1}$$

$$= \frac{\pi A_0^{(n-1)/2}}{\sqrt{2}\sqrt{n + 1}} \left(\int_0^1 \frac{du}{\sqrt{1 - u^{n+1}}} \right)^{-1} \tag{3.7-4a}$$

For values of n less than 3, the definite integral above can either be evaluated directly ($n = 1$) or expressed in terms of standard elliptic integrals ($n = 2,3$).   Values of n in excess of 3 can be accommodated by numerical integration.   The DF solution is, by virtue of Eq. (2.3-20),

$$\omega_{0(\text{DF})} = \sqrt{N(A_0)}$$

$$= \sqrt{\frac{2}{\sqrt{\pi}} \frac{\Gamma[(n + 2)/2]}{\Gamma[(n + 3)/2]}} A_0^{(n-1)/2} \tag{3.7-4b}$$

As both solutions for oscillation frequency display identical dependences upon A_0, the percent error attendant on DF utilization can be assessed with a single calculation.   Table 3.7-1 lists the result of this calculation for various nonlinearity powers from 1 ($y = x^1$,
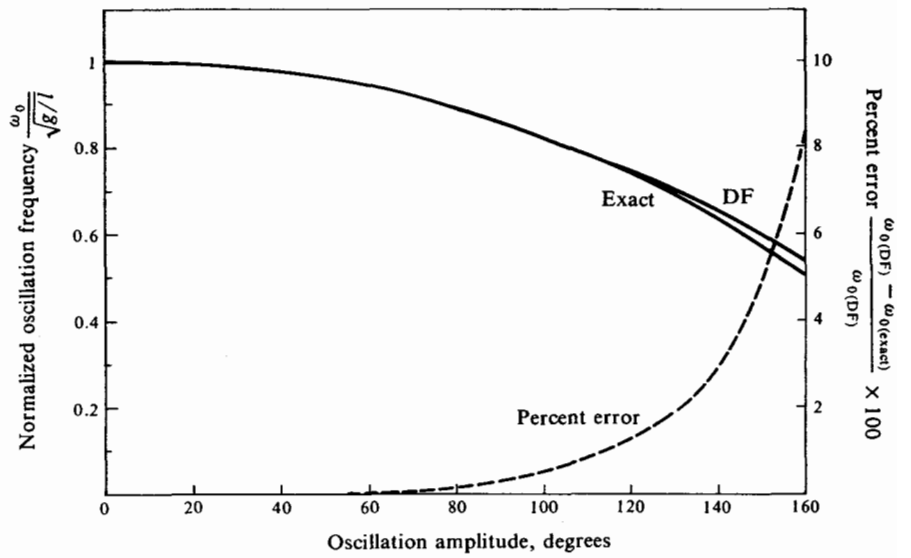
Figure 3.7-3 *Conservative-free-oscillation parameters for a simple pendulum.*

linear case) to 9 ($y = x^9$). Also tabulated is the ratio $|A_3/A_1|_x$ *at the input to the nonlinearity.* This ratio is computed by taking the harmonic amplitude ratio $|A_3/A_1|_y$ at the nonlinearity output (having assumed a sinusoidal input) and multiplying by $\frac{1}{9}$, the amount by which the linear element attenuates the third harmonic relative to the first harmonic.

A very significant, albeit quite expected, conclusion can be drawn from this tabulation, namely, that the DF solution accuracy improves with decreasing third-harmonic magnitudes

TABLE 3.7-1 FREE-OSCILLATION RESULTS IN A CONSERVATIVE SECOND-ORDER SYSTEM WITH AN ODD POLYNOMIAL NONLINEARITY

n	Describing function result $A_0^{(1-n)/2} \times \omega_{0(DF)}$	Exact result $A_0^{(1-n)/2} \times \omega_{0(exact)}$	Percent error $\dfrac{\omega_{0(DF)} - \omega_{0(exact)}}{\omega_{0(DF)}} \times 100\%$	Harmonic amplitude ratio at x $\left|\dfrac{A_3}{A_1}\right|_x \times 100\%$
1	1	1	0	0
2	0.921	0.914	0.76	2.2
3	0.866	0.847	2.2	3.7
4	0.824	0.795	3.5	4.8
5	0.791	0.748	5.3	5.6
6	0.763	0.711	6.7	6.2
7	0.740	0.678	8.4	6.7
8	0.719	0.648	10.0	7.1
9	0.702	0.622	11.3	7.4

at the nonlinearity input. For values of the ratio $|A_3/A_1|_x$ less than 5 percent, it is seen that DF results are accurate to better than 5 percent. This result is quite representative of DF accuracy in second- and higher-order systems.

Example 3.7-4[1]   Determine the limit cycle amplitude of a pulse-excited clock pendulum with damping ratio ζ. The velocity variation resulting from impact is a constant, V.

  The equation we are dealing with is

$$\ddot{x} + 2\zeta\dot{x} + x = f(x,\dot{x})$$

where the nonlinear characteristic is as shown in Fig. 3.7-4. By assuming a sinusoidal input, $x = A \sin \omega t$, and computing the DF, we find

$$n_p(A,\omega) = 0 \qquad n_q(A,\omega) = \frac{4D\delta}{\pi A^2}$$

Since $f(x,\dot{x})$ is a narrow pulse, it follows by integration of the differential equation of motion over the pulse duration that the change in velocity is equal to the pulse area. That is ($\tau = $ pulse-time duration),

$$V = \tau D$$

Continuing, the velocity of the assumed sinusoidal oscillation in the vicinity of $t = 0$ $[\dot{x}(0) = A_0\omega_0]$ times the pulse-time duration must equal 2δ, the nonlinearity "on" width.

$$2\delta = A_0\omega_0\tau$$

Combining these equations, we get for the linearized nonlinearity the DF representation [Eq. (2.2-14)]

$$f(x,\dot{x}) = n_p(A_0,\omega_0)x + \frac{n_q(A_0,\omega_0)}{\omega_0}\,\dot{x} = \frac{2V}{\pi A_0}\,\dot{x}$$

Inserting this into the original equation of motion yields

$$\ddot{x} + 2\left(\zeta - \frac{V}{\pi A_0}\right)\dot{x} + x = 0$$

from which it is clear that we get a steady-state oscillation only if

$$A_{0(DF)} = \frac{V}{\pi\zeta} \qquad\qquad (3.7\text{-}5a)$$

The subscript DF is used to distinguish this result from the exact limit cycle amplitude found by Magnus (*op. cit.*),

$$A_{0(exact)} = V\frac{\exp\left[-(\zeta/\sqrt{1-\zeta^2})\tan^{-1}(\sqrt{1-\zeta^2}/\zeta)\right]}{1 - \exp\left[-(\pi\zeta/\sqrt{1-\zeta^2})\right]} \qquad (3.7\text{-}5b)$$

Figure 3.7-5 is a plot of exact and approximate solutions as well as percent error in DF solution versus ζ. For low ζ the pendulum is nearly undamped and oscillates almost sinusoidally. The error in this region is small. The DF solution percent error is seen to

---

[1] This is one of many examples investigated by Magnus (Ref. 65) in an excellent paper on DF theory.
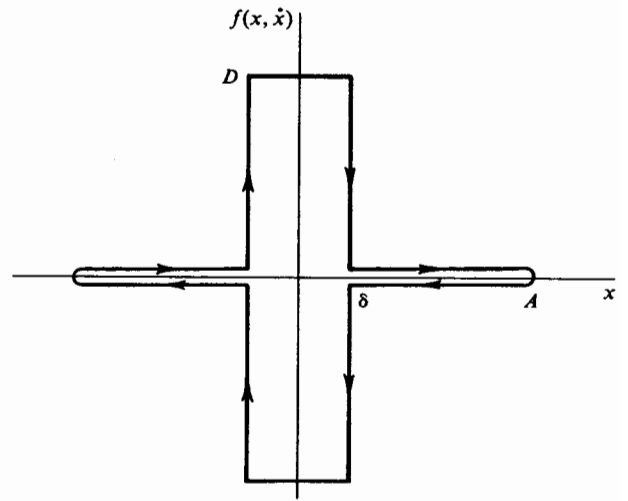
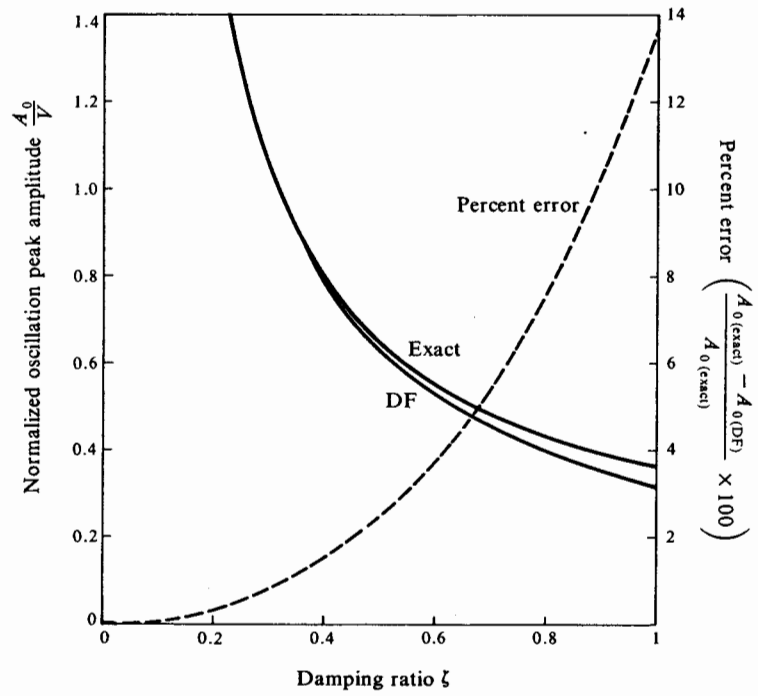**Figure 3.7-4** *Nonlinear characteristic of clock-pendulum pulse exciter.*

Figure 3.7-5 *Solution for limit cycle amplitude of a pulse-excited clock pendulum.*

increase with increasing ζ.   We note here that with increasing ζ the oscillation becomes more nonsinusoidal.   This is due to failure of the loop to attenuate higher harmonics sufficiently, relative to the fundamental.

Example 3.7-5   Investigate the limit cycle behavior of a temperature control system for which the block diagram is as shown in Fig. 3.7-6a.

Using DF methods, we immediately conclude that *no* oscillation is predicted to take place.   The graphical construction pertinent to this conclusion is shown in Fig. 3.7-6b. The DF result is obviously in complete error, since a limit cycle must occur in this on-off system if $DK > \delta$.   Why, then, we may ask, does the DF formulation yield such poor results?
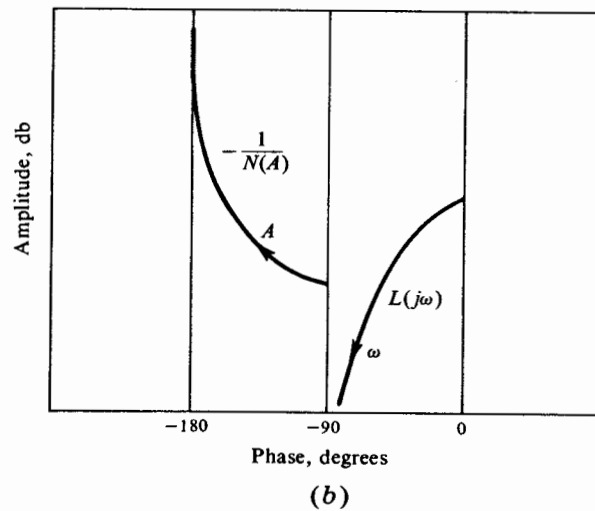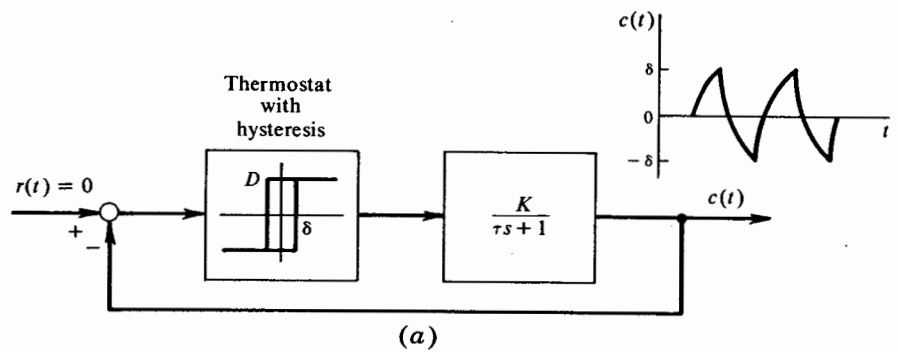

$$(a)$$

$$(b)$$

Figure 3.7-6   (a) *Temperature-regulating system.*   (b) *DF amplitude-phase-plot limit cycle construction.*

The answer to this enigma becomes evident if the exact limit cycle waveform is considered. This has been illustrated in Fig. 3.7-6a, and is seen to consist of matched single-exponential segments. The waveform suffers slope discontinuities twice during each limit cycle period, and is, generally speaking, hardly sinusoidal. *It is waveform harmonic content which is responsible for the DF failure.*

Without knowledge of the exact waveform, however, these poor results can be and should have been anticipated. A simple calculation demonstrates this point. The third- to first-harmonic amplitude ratio has been shown to be $\frac{1}{3}$ at the nonlinearity output. The linear elements attenuate each of the harmonics by the factor $K/\sqrt{1 + n^2\omega_0^2\tau^2}$, where n is the harmonic number. Thus, at the *input* to the nonlinearity, we have

$$\left|\frac{A_3}{A_1}\right|_x = \frac{1}{3}\sqrt{\frac{1 + \omega_0^2\tau^2}{1 + 9\omega_0^2\tau^2}} \qquad (3.7\text{-}6)$$

Without specifying the value of $\omega_0\tau$ it is clear that the following limiting inequality holds:

$$0.111 < \left|\frac{A_3}{A_1}\right|_x < 0.333 \qquad (3.7\text{-}7)$$

That is, at the input to the nonlinearity, the third-harmonic amplitude is somewhere between 11.1 percent and 33.3 percent of the fundamental. As a result of this large amplitude at x, the unaccounted for phase shift of the third harmonic is sufficient to upset the original input zero-crossing assumptions (in this *on-off relay* problem only the input zero crossings, *not* its detailed waveshape, are influential on limit cycle behavior), and thus negate the DF calculation.

That the DF approach fails in the last example is not of the slightest consequence; there are better ways to study this low-order system.[1] What is important is the fact that clear indication exists which shows the basic assumptions of DF theory to be unsatisfied. We shall pursue this point further.

IMPORTANCE OF HARMONICS (CONTINUED)

Reviewing each of the foregoing examples, we can make one consistent observation, namely, that DF solution accuracy was always degraded by an increasing harmonic content at the input to the nonlinearity.

It is not sufficient merely to know the harmonic content at the nonlinearity output; the filtering influence of the linear elements must also be assessed. From what we have seen so far, a doubly integrating linear filter provides sufficient attenuation of higher harmonics to yield very reasonable DF results

---

[1] Exact solution by matching exponential waveforms at the switching boundaries, for example. Another method of study, the phase plane (see Chap. 1), is of general use for the efficient and complete study of first- and second-order nonlinear systems. Still another method, Tsypkin's exact method, is of use in the study of limit cycles in relay systems and is examined in some detail later in this chapter. At that time Example 3.7-5 is treated in a most satisfactory manner.

for a wide variety of nonlinearities, whereas a first-order linear lag filter insufficiently attenuates these harmonics and can lead to incorrect DF results. Many other examples have been studied along the same lines. The results indicate clearly that as the order[1] of the linear portion of the system *increases*, the error in DF application is *reduced*. This is an interesting situation since the phase plane and other related techniques are well suited to the study of second- (and occasionally third-) order systems. In this light the DF ought to be considered as an adjunct to these methods, for use in studying third- (and occasionally second-) as well as higher-order systems. That higher-order systems are more amenable to simple DF analysis is indeed a happy circumstance, being quite the opposite of a trend one normally anticipates in system work. The reason for this is simply that repeated integrations of any periodic waveform of frequency ω eventually reduce that waveform to a sinusoid of frequency ω. Consider an arbitrary periodic function $y(t)$ expressed in its complex Fourier series representation.

$$y(t) = \sum_{k=-\infty}^{\infty} A_k e^{jk\omega t} \tag{3.7-8}$$

Let us integrate this expression m times and dispose of zero-frequency terms.

$$y^{(m)}(t) = \sum_{k=-\infty}^{\infty} \frac{A_k}{(jk\omega)^m} e^{jk\omega t}$$

$$= \sum_{k=-\infty}^{\infty} A_k^{(m)} e^{jk\omega t} \tag{3.7-9}$$

The superscript denotes the number of integrations performed. If we form the ratio of kth harmonic to first harmonic and let $m \rightarrow \infty$, we see indeed that only the first harmonic remains.

$$\frac{A_k^{(m)}}{A_1^{(m)}} = \frac{[A_k/(jk\omega)^m]}{[A_1/(j\omega)^m]} = \frac{1}{k^m} \frac{A_k}{A_1} \rightarrow 0 \qquad \text{as } m \rightarrow \infty \tag{3.7-10}$$

The study of repeated integrations on a square wave points out dramatically how the fundamental square-wave frequency is eventually singled out. Figure 3.7-7 is a demonstration of this process, with each successive integral normalized in amplitude and freed of the bias resulting from integration. After the second integration the (parabolic) waveform already looks quite sinusoidal. After an infinite number of integrations, a pure sinusoid is converged upon. We hasten to point out that this exercise is meant only to lend physical credence to the accuracy development. Replacement of the pure integrations by first-order-lag filters would tend to slow down the convergence to a sinusoid, etc.

[1] More specifically, the excess of poles over zeros in the frequency region up to $5\omega_0$. This presumes no lightly damped pole pairs at the frequency of nonlinearity output harmonics.
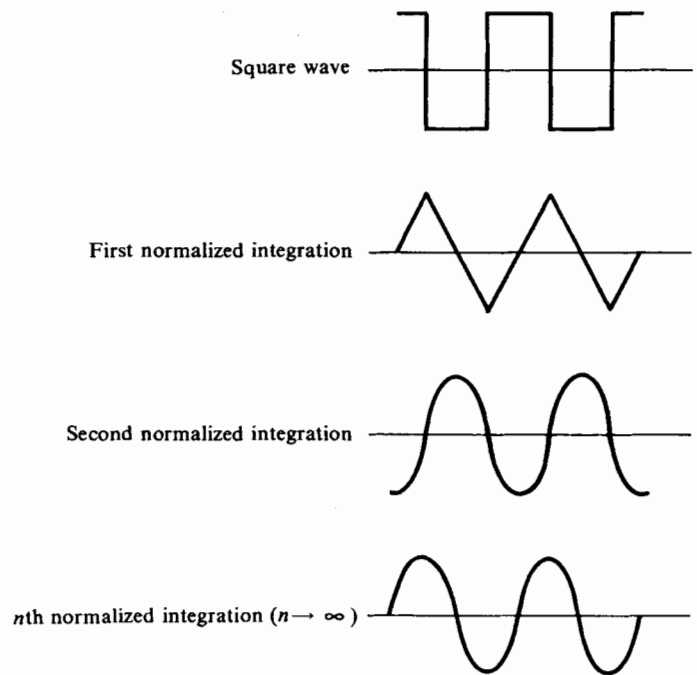
Square wave

First normalized integration

Second normalized integration

nth normalized integration $(n \rightarrow \infty)$

Figure 3.7-7 *Effect of repeated integration on a square wave (normalized).*

The use of asymptotic Bode plots is very convenient for determining the attenuation of all harmonic frequencies by the loop linear elements. Of course, the same information also exists on amplitude-phase and polar plots, but these plots are parametrized by ω, and hence must be carefully examined to ascertain actual attenuation ratios. Figure 3.7-8 shows the amplitude-phase plots for two different linear open-loop systems. Were the loops to be closed by the insertion of a non-phase-shifting nonlinearity, one ought to feel secure in his estimation that DF analysis would yield superior results in case *a* as opposed to case *b*, simply because of the observation that third- and higher-harmonic filtering is more pronounced in this case.

A word of caution with regard to the accuracy expected in graphical interpretation of DF solutions is in order. When the loci $L(j\omega)$ and $-1/N(A,\omega)$ indicated on the amplitude-phase plot are nearly orthogonal at the limit cycle intersection, results are ordinarily of 5 percent to 10 percent accuracy. Nearly parallel loci at intersection are apt to yield substantially poorer results. Low-frequency limit cycles often exist when the two loci approach each other at low frequencies but indicate no intersection. Such limit cycles can occasionally be determined by making a second-order
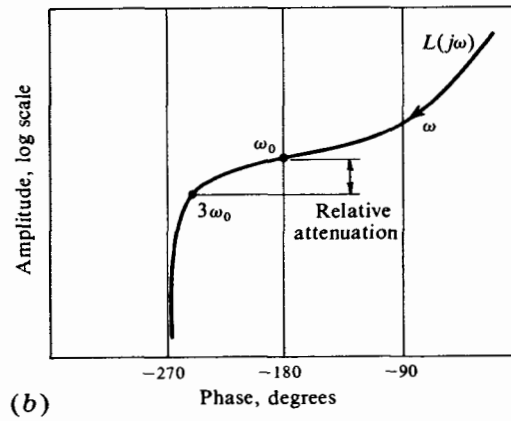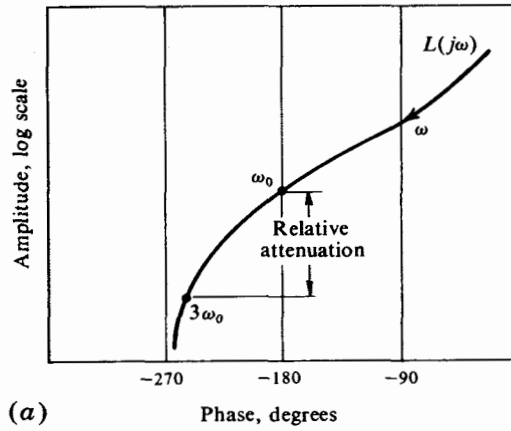
(a)

(b)

Figure 3.7-8 Estimation of the importance of harmonics by use of the amplitude-phase plane. (a) Less important; (b) more important.

approximation to the linear elements and studying the resultant system by phase-plane techniques. At any rate, in the latter instance it is desirable to determine a higher-order DF approximation as a check on DF accuracy. One such approximation is presented next. Another can be found in Popov (Ref. 88).

REFINED DF APPROXIMATION

Refining the DF formulation implies accounting for some of the higher harmonics comprising the residual, heretofore neglected. The means for achieving such DF refinement must be uncomplicated, for without this

constraint all the benefit of DF usage will be mitigated.   We shall proceed
with this point of view.

What we are after is essentially a new DF which describes the fundamental
gain of the nonlinearity in the presence of the fundamental and *simultaneously*
some of the higher harmonics.   In Chap. 5 we shall see that *exact* calcula-
tions for one such DF are indeed possible, but quite complicated.   Even
then only the third harmonic is considered.   Although such calculations
definitely have their own domain of utility (study of subharmonic responses,
for example), we should still like to seek out a DF second approximation,
whose primary function is to serve as an accuracy check in areas of question-
able DF results.

A consideration of the significance of feedback in linear systems provides
us with the direction to follow, in order to find a scheme for DF refinement.
Consider a unity feedback linear system with forward-path elements $L(j\omega)$
and input $R(j\omega)$.   The output $C(j\omega)$ can be constructed by tracing the input
around the loop an infinity of times.   Thus, when the input is first applied,
the output is, for an instant, $R(j\omega)L(j\omega)$.   Presently, some feedback occurs,
and the output is reduced by the amount $-R(j\omega)L^2(j\omega)$.   Additional feed-
back occurs, and the following series can be constructed ($|L(j\omega)| < 1$):

$$C(j\omega) = R(j\omega)L(j\omega) - R(j\omega)L^2(j\omega) + R(j\omega)L^3(j\omega) - \cdots$$

$$= R(j\omega)L(j\omega)[1 - L(j\omega) + L^2(j\omega) - \cdots]$$

$$= R(j\omega)\frac{L(j\omega)}{1 + L(j\omega)} \qquad (3.7\text{-}11)$$

This, of course, is the familiar linear-system closed-loop transfer function
result.   We now make the observation that if $L(j\omega)$ is a low-pass function,
higher powers of $L(j\omega)$ tend to zero in the frequency range of interest.
Under this condition the transfer function can be *approximated* by taking
only the first few terms of the series, namely,

$$\frac{C(j\omega)}{R(j\omega)} \approx L(j\omega)[1 - L(j\omega)] \qquad (3.7\text{-}12)$$

This process is tantamount to considering only the first portion, as it were,
of an infinite series of feedbacks.   The same approach has been adopted for
synthesizing a refined DF approximation (Ref. 34).

Suppose that by means of DF theory the nonlinear system of Fig. 3.7-9 has
been found to oscillate at a frequency ω.   In this first DF approximation,
the input to the nonlinearity, $x_{(1)}(t)$, has been taken as a pure sinusoid.   All
higher harmonics were ignored.   As a second approximation we now take
the input to the nonlinearity to be the pure sinusoid calculated above plus
the fed-back quantity comprised of the filtered residual.   We now compute
the output of the nonlinearity in response to the second-approximation

residual
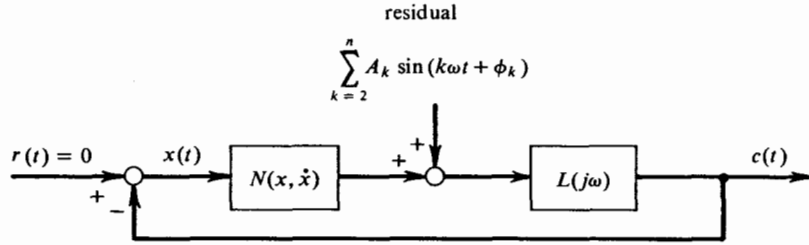
$$\sum_{k=2}^{n} A_k \sin(k\omega t + \phi_k)$$

First approximation: $x_{(1)}(t) = A \sin \omega t$

Second approximation: $x_{(2)}(t) = A \sin \omega t - \sum_{k=2}^{n} A_k |L(jk\omega)| \sin(k\omega t + \phi_k + \underline{/L(jk\omega)})$

Figure 3.7-9 *Formulation of a refined DF approximation.*

input $x_{(2)}(t)$, and define the second-approximation DF in the conventional manner as the complex ratio of output first-harmonic term to input first-harmonic term. The effect of higher harmonics is implicit in this formulation, although they have by no means been accounted for exactly. To obtain corrected results, one then uses the refined DF and performs otherwise ordinary DF calculations. The technique is demonstrated in detail by the following example.

Example 3.7-6 Determine the refined DF solution for the limit cycle displayed by the third-order ideal-relay control system of Fig. 3.7-10. Relay drive levels are $\pm D$.

The first DF solution for the limit cycle frequency and amplitude is easily found to be

$$\omega_0 = \omega_n \qquad A_0 = \frac{2DK}{\pi \zeta \omega_n}$$

Hence the first-approximation nonlinearity input is

$$x_{(1)}(t) = \frac{2DK}{\pi \zeta \omega_n} \sin \omega_n t$$

The residual consists of the odd harmonics associated with a square wave of amplitude D, as shown in Fig. 3.7-10. By passing this residual through $L(j\omega)$, we find as the second approximation to the nonlinearity input

$$x_{(2)}(t) = \frac{2DK}{\pi \zeta \omega_n} \sin \omega_n t - \sum_{k=3,5,\ldots}^{\infty} \frac{4D}{\pi k} |L(jk\omega_n)| \sin [k\omega_n t + \underline{/L(jk\omega_n)}]$$

where the amplitude and phase transfer of $L(jk\omega_n)$ are $(k > 1)$

$$|L(jk\omega_n)| = \frac{K\omega_n^2}{k\omega_n \sqrt{(\omega_n^2 - k^2\omega_n^2)^2 + (2k\zeta\omega_n^2)^2}}$$
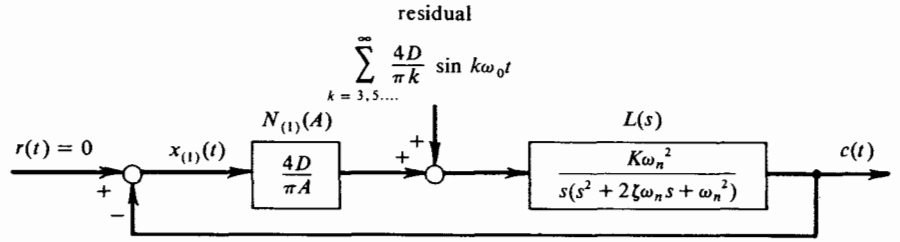
$$= \frac{K}{k\omega_n \sqrt{(k^2 - 1)^2 + (2k\zeta)^2}}$$

residual

$$\sum_{k=3,5,\ldots}^{\infty} \frac{4D}{\pi k} \sin k\omega_0 t$$

$r(t) = 0$ $x_{(1)}(t)$ $N_{(1)}(A)$ $\dfrac{4D}{\pi A}$ $L(s)$ $\dfrac{K\omega_n{}^2}{s(s^2 + 2\zeta\omega_n s + \omega_n{}^2)}$ $c(t)$

Figure 3.7-10 *Third-order ideal-relay control system, first DF approximation.*

and

$$\underline{/L(jk\omega_n)} = -\frac{\pi}{2} - \tan^{-1}\frac{2k\zeta\omega_n{}^2}{\omega_n{}^2 - k^2\omega_n{}^2}$$

$$= -\frac{3\pi}{2} + \tan^{-1}\frac{2k\zeta}{k^2 - 1}$$

Let us now restrict our attention to the third harmonic only. In principle, additional harmonics can be carried along, but for demonstration they serve only to clutter the issue. Thus $k = 3$, and $x_{(2)}(t)$ can be simplified to the following form:

$$x_{(2)}(t) = \frac{2DK}{\pi\zeta\omega_n}\left[\sin\omega_n t - \frac{\zeta}{36\sqrt{1 + (3\zeta/4)^2}}\cos\left(3\omega_n t + \tan^{-1}\frac{3\zeta}{4}\right)\right]$$

With this as the new input to the nonlinearity, the output will again be a square wave with fundamental-component amplitude $4D/\pi$. For this reason the *magnitude* of the refined DF is unchanged, remaining $4D/\pi A$. The *phase shift* of the refined DF is different from zero degrees, however, because of the presence of the third harmonic at the relay input. Presuming that the first DF approximation was not in great error, the sign of $x_{(2)}(t)$ will change at some value of t which is small ($t \approx 0$), allowing for an expansion of $x_{(2)}(t)$ in the vicinity of $t = 0$. Performing this expansion and setting $x_{(2)}(t)$ to zero yields

$$\omega_n t - \frac{\zeta}{36\sqrt{1 + (3\zeta/4)^2}}\left[\cos\left(\tan^{-1}\frac{3\zeta}{4}\right) - 3\omega_n t \sin\left(\tan^{-1}\frac{3\zeta}{4}\right)\right] = 0$$

Solving for $\omega_n t$, we find

$$\omega_n t = \frac{2\zeta}{9(8 + 5\zeta^2)}$$

Since the solution indicates a *positive* value for t, the phase shift associated with the refined DF is *lagging*. Thus the refined DF, designated $N_{(2)}(A)$, is

$$N_{(2)}(A) = \frac{4D}{\pi A}\exp\left[\frac{-j2\zeta}{9(8 + 5\zeta^2)}\right]$$

The refined approximate system for limit cycle study is illustrated in Fig. 3.7-11. It is impractical to observe the effect of $N_{(2)}(A)$ on the loop graphically, since the maximum phase shift involved is only $1.1°$ (at $\zeta = \sqrt{2}$). The condition for loop oscillation can be
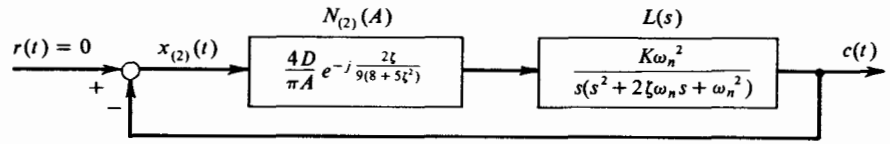
$$r(t) = 0 \qquad x_{(2)}(t)$$

$$N_{(2)}(A) \qquad\qquad L(s)$$

$$\frac{4D}{\pi A}\, e^{-j\frac{2\zeta}{9(8+5\zeta^2)}} \qquad\qquad \frac{K\omega_n{}^2}{s(s^2+2\zeta\omega_n s+\omega_n{}^2)} \qquad c(t)$$

Figure 3.7-11   *Third-order ideal-relay control system, refined DF approximation.*

studied analytically, however.   We require an open-loop phase shift of $-180°$, at the new limit cycle frequency ω_0; thus

$$-\frac{\pi}{2} - \tan^{-1}\left(\frac{2\zeta\omega_0\omega_n}{\omega_n{}^2 - \omega_0{}^2}\right) - \frac{2\zeta}{9(8+5\zeta^2)} = -\pi$$

or

$$\tan^{-1}\frac{2\zeta(\omega_0/\omega_n)}{1 - (\omega_0/\omega_n)^2} = \frac{\pi}{2} - \frac{2\zeta}{9(8+5\zeta^2)}$$

Continuing,

$$\frac{2\zeta(\omega_0/\omega_n)}{1 - (\omega_0/\omega_n)^2} = \tan\left[\frac{\pi}{2} - \frac{2\zeta}{9(8+5\zeta^2)}\right]$$

$$= \cot\frac{2\zeta}{9(8+5\zeta^2)}$$

$$\approx \frac{9(8+5\zeta^2)}{2\zeta}$$

Solving for ω_0, we finally obtain

$$\omega_0 \approx \omega_n\left[1 - \frac{2\zeta^2}{9(8+5\zeta^2)}\right]$$

To a second approximation we find that the limit cycle frequency does indeed depend upon ζ.   In the region of high ζ, the frequency correction term is largest.   Such an outcome is reasonable, since it is precisely in this region that the limit cycle waveform harmonic content is greatest.

   The first-order DF solution for ω_0, as well as the second-order and exact solutions, are illustrated in Fig. 3.7-12.[1]   The exact solution for the limit cycle frequency is the *implicit* relationship

$$\frac{\omega_n}{\omega_0} = \frac{2}{\pi\sqrt{1-\zeta^2}}\left[\frac{\sin\left(2\tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta}\right)\sinh\left(\frac{\pi\zeta\omega_n}{\omega_0}\right) - \cos\left(2\tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta}\right)\sin\left(\frac{\pi\sqrt{1-\zeta^2}\omega_n}{\omega_0}\right)}{\cosh\left(\frac{\pi\zeta\omega_n}{\omega_0}\right)+\cos\left(\frac{\pi\sqrt{1-\zeta^2}\omega_n}{\omega_0}\right)}\right]$$

which can be obtained by the method of piecewise solution of differential equations and boundary condition matching or Tsypkin's method (Sec. 3.8), among others.   As can be

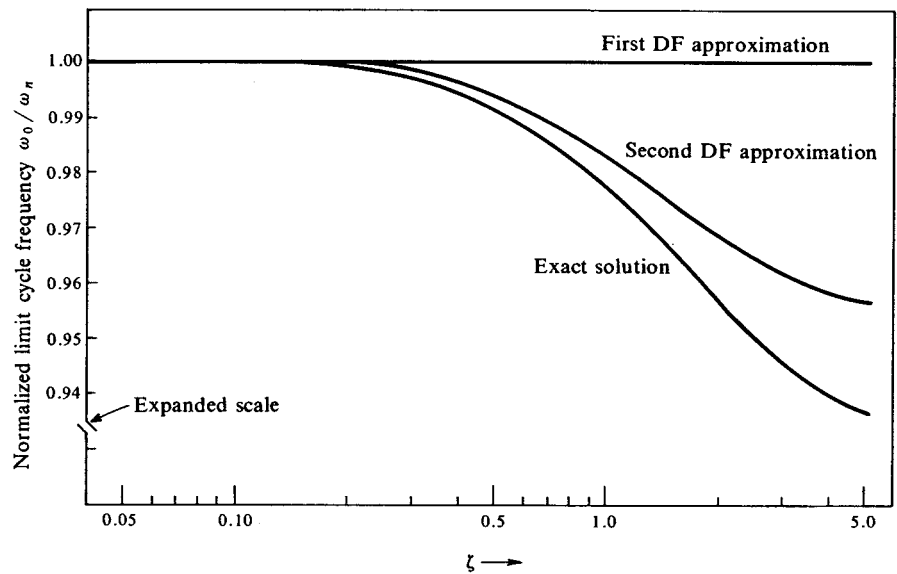[1] Note the expanded ordinate scale in this figure.

Figure 3.7-12   *DF and exact limit cycle frequency solutions for example third-order relay system.*

seen, it is a rather awkward expression, and is most reasonably dealt with by digital computer.

In the range $0 < \zeta < 1$, the first DF approximation has a maximum error of 2.3 percent, whereas the second DF approximation error is less than 0.6 percent.   Over the semi-infinite range in ζ, the first DF approximation is in maximum error of 10.3 percent ($\zeta \rightarrow \infty$), whereas the second DF approximation has a maximum error of 4.9 percent ($\zeta \rightarrow \infty$). From these data it is clear that the second approximation does yield qualitative results missed by the first approximation, namely, the ζ dependence of ω_0.

Note that in the region of high ζ, an amplitude-phase plot would show the $-1/N(A)$ and $L(j\omega)$ loci to be nearly parallel at intersection.   It was pointed out earlier that under this circumstance the accuracy of DF results is expected to be degraded somewhat.   The validity of this rule of thumb is redetermined consistently, as in the above example.

DF ACCURACY AND VALIDITY STUDIES

Up to this point we have attempted to both justify the validity of DF usage and assess DF accuracy by way of examples.   To a certain extent this procedure has led us to develop various rules of thumb which permit exploitation of the DF as a device for nonlinear system study.   However, as must certainly be the case, the rules of thumb are primarily stated *qualitatively*. There exist, on the other hand, several fairly general approaches to the accuracy and validity questions.   These, we shall see, lead to *quantitative* statements concerning accuracy and validity.

The systems to which we are likely to apply DF methods are commonly quite complex. Therefore it will probably be no surprise to learn that the quantitative statements which we shall be able to make in literal terms will be cumbersome and of little value in actual analysis and design. Nevertheless, such statements, and the theories from which they are derived, do form a concrete mathematical basis for DF usage. It is for this reason alone that they are important in the context of our work and worthy of mention.

In what follows, because of space limitations, we present only the rudiments of several important accuracy and validity studies (along with source references). The interested reader is encouraged to follow up this presentation, especially since each of these studies is of some interest in its own right.

Bogoliubov and Mitropolsky (Ref. 8) extended the mathematical foundations upon which the earlier work of Krylov and Bogoliubov (Sec. 2.1) was based. One important result of their work was a recursive method for obtaining an asymptotic series solution of the autonomous equation (μ is a small parameter)

$$\ddot{x} + \omega_0^2 x = \mu f(x,\dot{x}) \tag{3.7-13}$$

The series solution is asymptotic in the sense that the error of the nth approximation is proportional to the $(n + 1)$st power of the small parameter μ.

The form of the series solution is

$$x = A \cos \psi + \mu x^{(1)}(A,\psi) + \mu^2 x^{(2)}(A,\psi) + \cdots \tag{3.7-14}$$

where $\psi = \omega_0 t + \theta$, and $x^{(i)}$ are periodic functions of ψ, with period 2π. Further, A and ψ are required to satisfy the following differential equations:

$$\dot{A} = \mu A^{(1)}(A) + \mu^2 A^{(2)}(A) + \cdots \tag{3.7-15a}$$

$$\dot{\psi} = \omega_0 + \mu \Psi^{(1)}(A) + \mu^2 \Psi^{(2)}(A) + \cdots \tag{3.7-15b}$$

This formulation excludes the appearance of secular terms in the solution, which arise in the usual method of expansion in powers of a small parameter. Using Eqs. (3.7-14) and (3.7-15), one seeks a solution of Eq. (3.7-13). That is, one wants to find $x^{(i)}(A,\psi)$, $A^{(i)}(A)$, and $\Psi^{(i)}(A)$, which, simultaneously, lead to a solution of Eq. (3.7-13) to prescribed accuracy. The procedure to be followed is analogous to that used in the perturbation method (Sec. 2.1).

Bogoliubov and Mitropolsky (*op. cit.*) point out that only the first two or three terms of Eq. (3.7-14) can be computed in practice, because of the complexity of the formulas involved. Hence the practical applicability of the method is governed by the asymptotic properties of a few-term solution as $\mu \to 0$, rather than the actual convergence of Eq. (3.7-14). It is readily shown that the first series term in this recursive solution is the same as that originally developed by Krylov and Bogoliubov.

Among many examples considered by Bogoliubov and Mitropolsky is the simple pendulum, for which we have already obtained DF and exact solutions (Example 3.7-2). The DF solution, of course, is identical with the first-approximation solution of Bogoliubov and Mitropolsky. This follows since we have already indicated the equivalence of their solution to that of Krylov and Bogoliubov, and have earlier shown (Sec. 2.2) the relationship of the Krylov-Bogoliubov method to the DF method. For an oscillation amplitude of 160° we have seen that the first approximation to the oscillation frequency is in error by about 8 percent (Fig. 3.7-3). The second approximation, including the term $x^{(1)}(A,\psi)$, is reported (Ref. 8, p. 64) to be in error by only 3 percent. Considering the degree of nonlinearity present at such a large oscillation amplitude, we should conclude that these results are excellent and, in particular, that the second-order asymptotic solution can indeed represent a substantial improvement over the first-order solution. Unfortunately, we note that these results are applicable only to second-order systems.

Johnson (Ref. 48) studied the problem of DF accuracy directly. His work is based on earlier studies of Bulgakov (Ref. 12), and is addressed to the determination of the second term in a series for which DF analysis provides the first term. This additional term, once found, is used to correct DF results and to indicate DF solution accuracy. As we have already noted, general calculations of this sort are complicated and seldom of use in practice. Since this turns out to be the case here, we indicate Johnson's principal results without discussing his method in any detail.

Consider a single nonlinearity system to be in an oscillatory state. For a related hypothetical system define the input amplitude to the nonlinearity x and the oscillation frequency ω in terms of the following series:

$$x = x_0 + \mu x_1 + \frac{\mu^2}{2!} x_2 + \frac{\mu^3}{3!} x_3 + \cdots \qquad (3.7\text{-}16a)$$

$$\omega = \omega_0 \left(1 + \mu \lambda_1 + \frac{\mu^2}{2!} \lambda_2 + \frac{\mu^3}{3!} \lambda_3 + \cdots \right) \qquad (3.7\text{-}16b)$$

where μ is an artificially introduced parameter multiplying the nonlinearity, for which the value $\mu = 1$ reduces the hypothetical system to the original system. In Eqs. (3.7-16), called the *generating solution*, x_0 and ω_0 are the DF solutions to the original system; $x_i(i \geq 1)$ and $\lambda_i(i \geq 1)$ constitute the DF correction terms.

One of Johnson's main results was the determination that $\lambda_1 = 0$; i.e., the first frequency correction is zero. He also provided formulas for the first amplitude correction and the second frequency correction, although the accuracy estimates they provide depend upon the (unknown) radius of convergence of his power series. Bass (Ref. 3) states, appropriately, that

"Johnson's real accomplishment was to produce a new heuristic motivation for the [DF] method; one seeks for the linearized differential equation a linear equation possessing a 'generating solution' whose first frequency correction vanishes when it is perturbed in the 'direction' of the given nonlinear equation."

In his paper (*op. cit.*), Johnson studied a second-order system with friction-controlled backlash, for which the DF method results in limit cycle amplitude and frequency errors of 2.50 and 0.97 percent, respectively. Applying his second-frequency and first-amplitude corrections, he was able to reduce these errors to 0.28 and 0.23 percent, respectively. Similar order-of-magnitude reductions in DF errors were obtained in second-order ideal-relay- and ideal-velocity-limited servos. These results are outstanding, particularly in view of the fact that the linear elements' transfer function is only of second order. They not only tend to bear out the validity of Johnson's work but also provide further evidence that DF accuracy is on the order of 5 to 10 percent for a rather wide class of nonlinear systems.

Bass (Ref. 5) attempts to give a rigorous treatment of the mathematical validity of the describing function method by use of topological arguments. He starts with a physical system whose behavior is governed by a nonlinear *vector* differential equation of the form

$$\ddot{x} = f(x,\dot{x}) \qquad \text{where } f(-x, -\dot{x}) = -f(x,\dot{x}) \qquad (3.7\text{-}17)$$

Then, with the definition $\theta = \nu t$ ($\nu = 2\pi/T$, T being a period, if one exists; he is essentially normalizing the problem), Eq. (3.7-17) becomes

$$\nu^2 \frac{d^2x}{d\theta^2} = f\left(x, \nu \frac{dx}{d\theta}\right)$$

which is then broken into its linear and nonlinear parts to give the final form of the physical-system equation.

$$\nu^2 \frac{d^2x}{d\theta^2} + \nu C_1 \frac{dx}{d\theta} + C_2 x = \Phi\left(x, \nu \frac{dx}{d\theta}\right) \qquad (3.7\text{-}18)$$

Proceeding along typical DF technique lines, he defines the vector $x_0(\theta)$, which is just the fundamental harmonic term in the Fourier series of $x(\theta)$, and obtains what he terms the "hypothetical tractable system" equation

$$\nu^2 \frac{d^2x}{d\theta^2} + \nu C_1 \frac{dx}{d\theta} + C_2 x = \Phi\left(x_0, \nu \frac{dx_0}{d\theta}\right) \qquad (3.7\text{-}19)$$

The "describing function problem" is then to determine the circumstances under which, if Eq. (3.7-19) has precisely one periodic solution, the physical system of Eq. (3.7-18) will also have at least one periodic solution.

What Bass does show is that if $\Phi(x,\dot{x})$ is *smooth*, then under appropriate conditions a periodic solution to Eq. (3.7-19) implies a periodic solution to Eq. (3.7-18). If $\Phi(x,\dot{x})$ is piecewise-smooth, he is able only to show the conditions under which there are periodic solutions to Eq. (3.7-19), and not what this implies in respect to the physical system, Eq. (3.7-18). For $\Phi(x,\dot{x})$ piecewise-smooth, he shows that Eq. (3.7-19) has the appropriate periodic solution if and only if a $2n$-dimensional (x is an n vector) system of implicit equations is satisfied. The necessity of this condition is derived from consideration of the Fourier coefficients of $x(\theta)$ quite easily, though the sufficiency involves more work.

It might be remarked that Bass does not expect an engineer to verify theoretically all the requirements he states for his results to hold; that would be a very large and difficult undertaking. He is attempting to put the technique on a mathematically justifiable basis, not to give methods for calculations. Unfortunately, the requirement that $\Phi(x,\dot{x})$ [and consequently $f(x,\dot{x})$] be smooth leaves out a large class of important nonlinear control problems; however, he does supply these rigorous results for a "smooth forcing function" system. An important result of this work is a rigorous verification of the *filter hypothesis* of conventional DF usage, at least for smooth systems.

Sandberg (Ref. 91) considers the operator equation

$$x = LN(x + r) \tag{3.7-20}$$

where $x = $ a vector function,
  $L = $ a linear operator,
  $N = $ a nonlinear operator,
  $r = $ a periodic forcing function,

and the equivalent linearization approximation

$$x_0 = \tilde{L}N(x_0 + r) \tag{3.7-21}$$

in which $\tilde{L}$ extracts the fundamental component of the Fourier series expansion of $LN(x + r)$ and x_0 is a periodic solution. Sandberg's analysis is carried out in the space of periodic functions square integrable over a period. His method is to determine conditions that guarantee that an operator derived from LN is a contraction mapping in the whole space. He presents conditions under which there exists a unique periodic response to an arbitrary periodic input with the same period, as well as an upper bound on the mean-squared error in using equivalent linearization. He also gives conditions under which subharmonics and self-sustained oscillations cannot occur.

Holtzman (Ref. 45) treats essentially the same problem as Sandberg (*op. cit.*). His approach differs from Sandberg's in that Holtzman tries to obtain

a contraction mapping only in a neighborhood of x_0, and thus avoids global Lipschitzian type of requirements which often limit the applicability of the contraction mapping theorem. However, his analysis is not applicable to piecewise-linear nonlinearities as a result of requiring differentiability. In his work there are applications to time-lag systems and subharmonic oscillators.

In a different study, Holtzman compares the jump resonance criteria given by the DF method and the *circle criterion*, the latter being a mathematically rigorous sufficient condition for the absence of jump phenomena in nonlinear feedback systems. The result of that study, stated briefly, is that the circle criterion does not contradict the DF criterion. This can be construed as providing another favorable piece of evidence relative to validity of the DF method.

V. M. Popov (Ref. 89) studied the stability of a class of closed-loop nonlinear systems by using direct analytical methods. He succeeded in establishing certain general results which can be interpreted in terms of the system linear elements' frequency response and the loosely defined shape of the nonlinear characteristic. It is of considerable interest to examine DF analysis in view of these results.

The class of systems considered by Popov contains a single static memoryless nonlinearity which is sufficiently smooth to ensure the existence of a unique solution of the governing differential equations, and a linear part with only lagging phase. In one special case his results can be interpreted as follows: For nonlinear characteristics lying entirely within the first and third quadrants (shaded area of Fig. 3.7-13a), the system is stable provided only that the linear-part phase shift is always lagging and takes values between 0 and $-180°$.
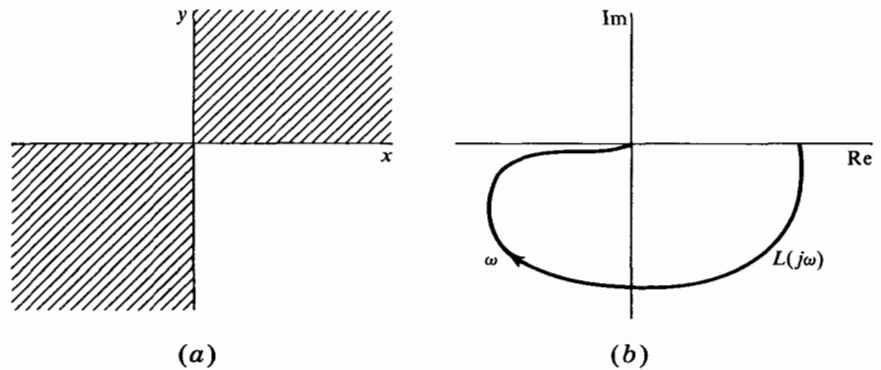

(a)   (b)

Figure 3.7-13   *Region of allowable nonlinear characteristics (a) and typical linear-part polar plot (b) for stability, according to V. M. Popov's criterion.*

Consider a DF analysis of this class of systems. As the nonlinear characteristic is single-valued and bounded by the lines $x = 0$ and $y = 0$, it follows (cf. Sec. 2.6) that $N(A)$ is real and of magnitude $0 \leq N(A) < \infty$. The polar plot of $-1/N(A)$ is thus the negative real axis. We immediately conclude that for all linear elements with phase lag never exceeding $-180°$ the nonlinear systems are stable, that is, no limit cycle takes place. This is in complete agreement with Popov's result, although, according to DF analysis, no restriction on allowable phase *lead* is required (except that it must be less than 180°).

A generalization of Popov's criterion for lagging phase shift has been accomplished in the determination of stability conditions when the nonlinear characteristic lies in the sector between the lines $x = 0$ and $x = Ky$ (Ref. 90), and this work has been extended to certain cases, including leading phase shift, by further restricting the shape of the nonlinear characteristic (Ref. 115). Since there presently exists no neat polar-plane interpretations of these results, they are not further pursued here. One does observe from such work, however, the hint that DF analysis may be in error when applied to systems in which large phase lead occurs at low frequencies. Experimental results[1] appear to verify this point.

It is further worth noting that, because of their generality, results based upon extensions of Popov's criterion are always more conservative than the corresponding DF statements.

3.8 EXACT METHODS FOR RELAY CONTROL SYSTEMS

A control system in which the power amplifier is a relay device is called a relay (alternatively, on-off, contactor, bang-bang, bistable, etc.) control system. The relay amplifier is desirable because it can be simple, rugged, compact, and relatively cheap, while meeting high load-power requirements. Examples of symmetrical characteristics of four relay amplifiers are shown in Fig. 3.8-1, where δ_h and δ_d indicate hysteresis and dead-zone switching levels, respectively, and the associated drive levels are $\pm D$. The relay as a control-system component is enjoying an increasing popularity in both old and new control problems. Home temperature control systems, aircraft and missile adaptive control systems, and space-vehicle attitude control systems constitute typical applications. We have already seen the use of a relay for time-optimal control (Sec. 3.4). The range of variety in relay controller utilization is, in fact, enormous.

The actual design of relay control systems is a matter elsewhere discussed

---

[1] Brought to the authors' attention by R. W. Brockett, Massachusetts Institute of Technology. See Prob. 5-20.
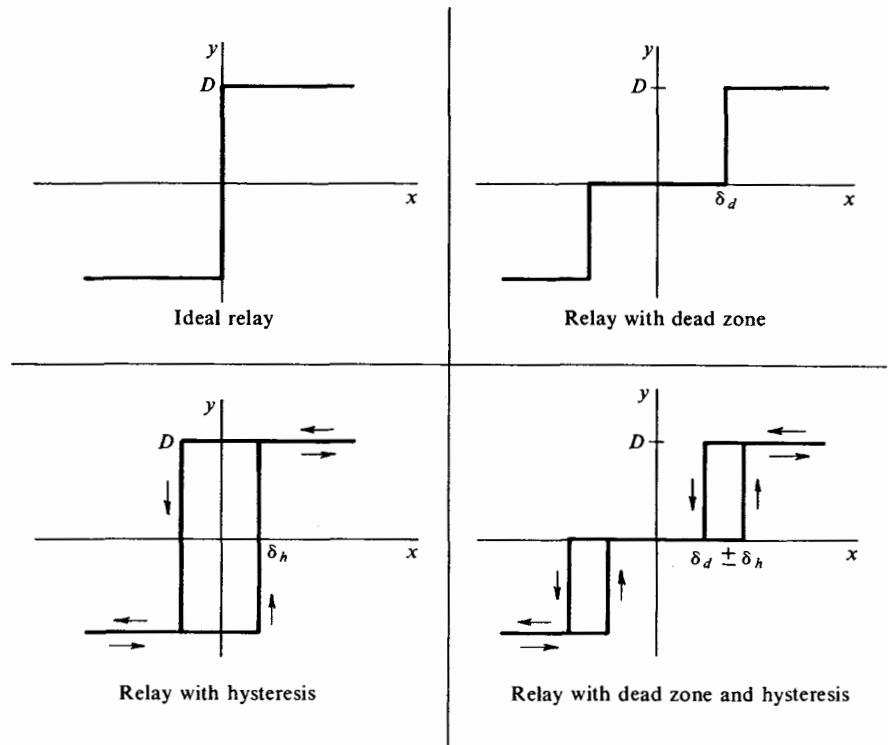
Figure 3.8-1   *Common relay characteristics.*

in this and other texts.   Of importance here is the fact that relay control systems frequently either sustain limit cycles, are required to operate in the non-limit-cycle mode near a region of possible limit cycle behavior, or (less frequently) are subjected to near-harmonic forcing.   Precise knowledge of steady oscillatory behavior is therefore desirable.   It turns out that for systems incorporating relay-type devices, exact methods for steady-oscillation determination are available.   Tsypkin's method is presented in some detail, because of its utility as a rapid check on DF results.   Inasmuch as this method is of limited value in actual control-system design, it is treated as a DF validity test.   As such, some discussion here is devoted to its analytical application, as opposed to its conventional graphical application.   The emphasis in presentation is on application to low-order systems since we are perhaps at this time convinced of the reliability of DF results as applied to high-order systems.   Although the development to follow deals only with symmetric two- or three-output-level devices, it is readily extended to asymmetric and other multiple-output-level devices.

PERIODICITY CONDITIONS

Consider a steady-state situation in which the closed-loop system of Fig. 3.8-2 sustains an oscillation. Sufficient conditions for the existence of this oscillation can be deduced, quite generally, from an observation of the periodic requirements on $x(t)$. For convenience we set a time origin at the instant of a positive switching (y becomes $+D$) and call T_0 the oscillation period ($T_0 = 2\pi/\omega_0$).

In the absence of any input $r(t)$, the oscillation will be symmetric about the origin for all symmetric relay characteristics. Thus conditions relating to periodicity need be written only over the half-cycle.

1. *Ideal relay.* The conditions sufficient for a periodic waveform are, in this case,[1]

$$x\left(\frac{T_0^-}{2}\right) = 0 \quad \text{and} \quad \frac{dx}{dt}\left(\frac{T_0^-}{2}\right) < 0 \qquad (3.8\text{-}1)$$

since switching takes place at the zero crossing point for x. Equation (3.8-1), geometrically motivated, merely states that the value of x at the half-period equals the value of x at the period outset and that the slope is reversed at these instants. These conditions follow from the assumed "mirror-image" symmetry, familiar from Fourier series study.

2. *Hysteretic relay.* As in the case of the ideal relay, there are only two switchings per cycle. Sufficient conditions for periodicity are

$$x\left(\frac{T_0^-}{2}\right) = -\delta_h \quad \text{and} \quad \frac{dx}{dt}\left(\frac{T_0^-}{2}\right) < 0 \qquad (3.8\text{-}2)$$

which can be argued in a manner similar to that presented above.

3. *Relay with dead zone.* For this nonlinearity there are four switchings per limit cycle period, or two over the half-period. The first switching occurs some time before the half-period, denoted $\rho T_0/2$, with $0 < \rho < 1$. The second switching occurs at the half-period. Both sets of conditions, which must simultaneously be met for possible oscillatory behavior, are

$$x\left(\frac{\rho T_0^-}{2}\right) = \delta_d \quad \text{and} \quad \frac{dx}{dt}\left(\frac{\rho T_0^-}{2}\right) < 0 \qquad (3.8\text{-}3a)$$

also

$$x\left(\frac{T_0^-}{2}\right) = -\delta_d \quad \text{and} \quad \frac{dx}{dt}\left(\frac{T_0^-}{2}\right) < 0 \qquad (3.8\text{-}3b)$$

[1] $\dfrac{T_0^-}{2} = \lim\limits_{\epsilon \to 0} \left(\dfrac{T_0}{2} - \epsilon\right)$

4. *Hysteretic relay with dead zone.* Again there are two switchings per half-period. The sufficient conditions for oscillatory behavior are

$$x\left(\frac{\rho T_0^-}{2}\right) = \delta_d - \delta_h \quad \text{and} \quad \frac{dx}{dt}\left(\frac{\rho T_0^-}{2}\right) < 0 \quad (3.8\text{-}4a)$$

also

$$x\left(\frac{T_0^-}{2}\right) = -\delta_d - \delta_h \quad \text{and} \quad \frac{dx}{dt}\left(\frac{T_0^-}{2}\right) < 0 \quad (3.8\text{-}4b)$$

where, as before, $0 < \rho < 1$.

Having noted the geometric prerequisites for oscillation sustenance, it is now possible to establish conditions on the linear loop elements $L(s)$, so as to satisfy these prerequisites. One method of accomplishing this task, introduced by Tsypkin, is discussed next.

TSYPKIN'S METHOD

Assume the system of Fig. 3.8-2 to be in a steady oscillatory mode. For the sake of discussion, assume the relay to have drive levels $\pm D$ and switching levels $\pm \delta_h$. $L(s)$ is assumed rational in s, having all its poles in the left half-plane, with at least as many poles as zeros. When oscillating, the relay is assumed to flip every half-cycle. Then, for any oscillation waveform, the nonlinearity output $y(t)$ will be a square wave.[1]

$$y(t) = Du(t) - 2Du(t - \tfrac{1}{2}T_0) + 2Du(t - T_0) - 2Du(t - \tfrac{3}{2}T_0) + \cdots$$
$$(3.8\text{-}5)$$

The Fourier series belonging to this periodic function is, for $t > 0$,

$$y(t) = \frac{4D}{\pi}\left(\sin \omega_0 t + \frac{1}{3}\sin 3\omega_0 t + \cdots + \frac{1}{k}\sin k\omega_0 t + \cdots\right)$$

$$= \frac{4D}{\pi}\sum_{k\,\text{odd}}^{\infty}\frac{\sin k\omega_0 t}{k}$$

$$= \frac{4D}{\pi}\sum_{k\,\text{odd}}^{\infty}\frac{1}{k}\,\text{Im}\,(e^{jk\omega_0 t}) \quad (3.8\text{-}6)$$

Following Bergen (Ref. 7), we note that since $c(t)$ and $\dot{c}(t)$ are not necessarily continuous, it is convenient to define a function $L_1(s)$ by

$$L(s) = L_1(s) + L(\infty) \quad (3.8\text{-}7)$$

---

[1] $u(t - \tau)$ is the displaced unit-step function of value 1 for $t \geq \tau$, 0 for $t < \tau$.
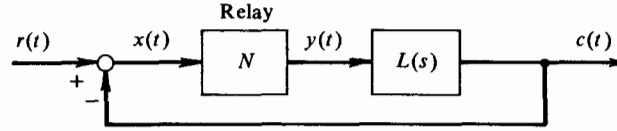
Figure 3.8-2 *Closed-loop nonlinear system.*

where $L_1(s)$ necessarily has more poles than zeros, and $L(\infty)$ is a constant. Since $L(s)$ has all its poles in the left half-plane, it follows from Eqs. (3.8-6) and (3.8-7) that

$$c(t) = \frac{4D}{\pi} \sum_{k \text{ odd}}^{\infty} \frac{1}{k} \text{Im} [L_1(jk\omega_0)e^{jk\omega_0 t}] + L(\infty)y(t) \qquad (3.8\text{-}8)$$

With $L_1(s)$ low-pass, the discontinuities in $c(t)$ appear only in the second term.

Limit cycles Since there is no loop input, $x(t) = -c(t)$, and the position periodicity condition [first part of Eq. (3.8-2)] at $t = T_0^-/2$ leads to

$$\sum_{k \text{ odd}}^{\infty} \frac{1}{k} \text{Im} [L_1(jk\omega_0)] = \frac{\pi}{4}\left[L(\infty) - \frac{\delta_h}{D} \right] \qquad (3.8\text{-}9)$$

For the velocity periodicity condition we require $\dot{c}(t)$, which is derived by differentiating Eq. (3.8-8) with respect to time.

$$\dot{c}(t) = \frac{4D\omega_0}{\pi} \sum_{k \text{ odd}}^{\infty} \text{Re} [L_1(jk\omega_0)e^{jk\omega_0 t}] + L(\infty)\dot{y}(t) \qquad (3.8\text{-}10)$$

In evaluating $\dot{c}(t)$ care must be taken to account for possible discontinuities contributed by $L_1(s)$. [The delta function due to $\dot{y}(t)$ occurs after the relay action and may be discarded.] Discontinuities will appear if

$$\lim_{s \to \infty} sL_1(s)$$

is nonzero, and the magnitude of each discontinuity is

$$\dot{c}\left(\frac{T_0^+}{2}\right) - \dot{c}\left(\frac{T_0^-}{2}\right) = -2D \lim_{s \to \infty} sL_1(s) \qquad (3.8\text{-}11)$$

The Fourier series of a piecewise-continuous function converges to the midpoint of each finite discontinuity; hence

$$\dot{c}\left(\frac{T_0^-}{2}\right) = -\frac{4D\omega_0}{\pi} \sum_{k \text{ odd}}^{\infty} \text{Re} [L_1(jk\omega_0)] + D \lim_{s \to \infty} sL_1(s) \qquad (3.8\text{-}12)$$

Noting that $\dot{x}(t) = -\dot{c}(t)$ in the absence of an input, the velocity periodicity condition [second part of Eq. (3.8-2)] becomes

$$\sum_{k\,\text{odd}}^{\infty} \text{Re } [L_1(jk\omega_0)] < \frac{\pi}{4\omega_0} \lim_{s \to \infty} [sL_1(s)] \qquad (3.8\text{-}13)$$

The periodicity equations derived above may be summarized by defining the so-called Tsypkin locus, $T(j\omega)$, as

$$T(j\omega) = \sum_{k\,\text{odd}}^{\infty} \text{Re } [L_1(jk\omega)] + j \sum_{k\,\text{odd}}^{\infty} \frac{1}{k} \text{Im } [L_1(jk\omega)] \qquad (3.8\text{-}14)$$

The Tsypkin locus is a plot of $T(j\omega)$ as ω is varied over the range zero to infinity. We may restate the necessary conditions for a symmetric limit cycle oscillation at some frequency ω_0 in terms of the Tsypkin locus, as follows:

$$\text{Im } [T(j\omega_0)] = \frac{\pi}{4}\left[L(\infty) - \frac{\delta_h}{D}\right]$$

$$\text{Re } [T(j\omega_0)] < \frac{\pi}{4\omega_0} \lim_{s \to \infty} [sL_1(s)] \qquad (3.8\text{-}15)$$

In the case for which $L(s)$ has at least two more poles than zeros, Eqs. (3.8-15) reduce to the more common form of the Tsypkin conditions

$$\text{Im } [T(j\omega_0)] = -\frac{\pi}{4}\frac{\delta_h}{D}$$

$$\text{Re } [T(j\omega_0)] < 0 \qquad (3.8\text{-}16)$$

Forced oscillations   Let us assume an input of the form

$$r(t) = R \sin (\omega_r t + \varphi)$$

and a corresponding loop *forced* oscillation at the frequency ω_r. Then $x(t) = r(t) - c(t)$, and the position and velocity periodicity conditions, derived in a manner completely analogous to that presented above, are

$$-R \sin \varphi + \sum_{k\,\text{odd}}^{\infty} \frac{1}{k} \text{Im } [L_1(jk\omega_r)] = \frac{\pi}{4}\left[L(\infty) - \frac{\delta_h}{D}\right]$$

$$-R \cos \varphi + \sum_{k\,\text{odd}}^{\infty} \text{Re } [L_1(jk\omega_r)] < \frac{\pi}{4\omega_r} \lim_{s \to \infty} [sL_1(s)] \qquad (3.8\text{-}17)$$

These conditions can be rewritten in terms of the Tsypkin locus $T(j\omega)$ as follows:

$$-R \sin \varphi + \text{Im } [T(j\omega_r)] = \frac{\pi}{4}\left[L(\infty) - \frac{\delta_h}{D}\right]$$

$$-R \cos \varphi + \text{Re } [T(j\omega_r)] < \frac{\pi}{4\omega_r} \lim_{s \to \infty} [sL_1(s)] \qquad (3.8\text{-}18)$$

Stability of periodic oscillations We have thus far been successful in determining whether a relay control system is *mathematically* capable of sustaining a periodic oscillation. The question now relates to whether in the presence of a small perturbation any given oscillating system will return to its original oscillatory state. The answer to this question will provide us with knowledge of the local asymptotic stability of an oscillatory mode in a *physical* relay system. Tsypkin (Ref. 106) describes a simple stability test based on the theory of sampled-data systems, the results of which we restate here:[1]

> A limit cycle mode in a relay control system is stable if the quantity $d\{\text{Im } [T(j\omega_0)]\}/d\omega_0$ is positive, and conversely, unstable if that quantity is negative.

The sign can be obtained either directly by inspection of the Tsypkin locus or analytically.

> A forced oscillatory mode in a relay control system is stable if the operating point lies to the left of $T(j\omega_r)$, and conversely, unstable if it lies to the right.

Stability in this case is determined graphically.

GRAPHICAL APPLICATION OF TSYPKIN'S METHOD

In order to apply this method graphically, one must construct the Tsypkin locus $T(j\omega)$, defined by Eq. (3.8-14). This construction can proceed either from the amplitude-phase-plane plot of $L_1(j\omega)$ with the aid of an overlay (Ref. 80) or directly from the polar plot of $L_1(j\omega)$. The latter alternative is rather easily executed, making the polar plot a particularly convenient representation for this method (see Fig. 3.8-3).

As a result of the low-pass nature of $L_1(j\omega)$, the Tsypkin locus coincides with the polar plot of $L_1(j\omega)$ at high frequencies. This decreasing importance of higher harmonics is the basis of first-harmonic linearization, and leads to the generalization that *DF limit cycle prediction tends to be more reliable in the higher limit cycle frequency region.* Just what constitutes a *higher* frequency region of course depends upon the particular linear elements in question. We might also observe that although any linear elements' polar plot and its corresponding Tsypkin locus may tend toward the same trajectory in the polar plane, the *frequency calibration* for each will necessarily be different (Prob. 3-19).

Graphical construction for study of limit cycles follows from Eqs. (3.8-15). The study of forced oscillations is also readily accomplished graphically. Recalling that Eqs. (3.8-18) must be simultaneously satisfied in order for a

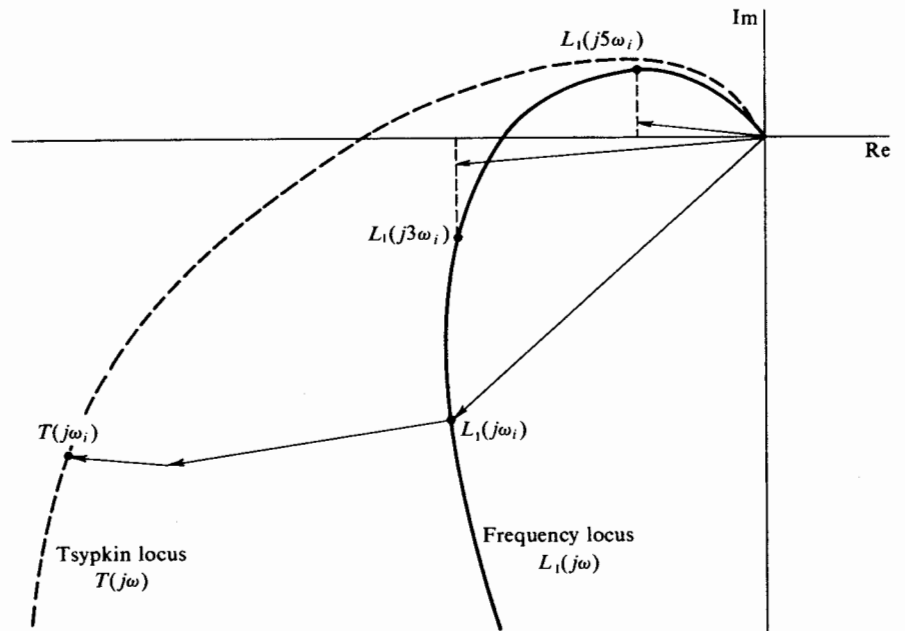[1] The derivation may be found in Gille et al. (Ref. 30, Chap. 26).

Figure 3.8-3 *Obtaining the Tsypkin locus from the transfer locus $L_1(j\omega)$. $T(j\omega_i)$ is computed from $L_1(j\omega_i)$, $L_1(j3\omega_i)$, $L_1(j5\omega_i)$ according to Eq. (3.8-14). (Adapted from Gille, Pélegrin, and Decaulne, Ref. 30.)*

forced oscillation to be possible, one is led to the construction of Fig. 3.8-4. By drawing a circle of radius R centered at the point on the Tsypkin locus given by $\omega = \omega_r$, all mathematical solutions are determined. In the illustration we see that two solutions of different phase appear possible. Based upon the stability test, however, we see that the solution labeled r_1 (to the left of ω_r) is the only stable solution, and that r_2 is in fact physically unrealizable. One may present some justification for the above argument at this point by noting that as $R \to \infty$, $\varphi_1 \to 0$, whereas $\varphi_2 \to \pi$. Certainly, we should expect the output to be in synchronism with the input in this limiting case, as the first solution indicates.

This graphical construction points out an interesting forced response pattern common to nonlinear systems, namely, *regional synchronous behavior*. As is evident from Fig. 3.8-4*b*, at a given input frequency ω_r for all input amplitudes below a certain value $R_{\min}$, no solution of Eqs. (3.8-18) is possible. The system actually sustains neither a limit cycle nor an oscillation at the frequency ω_r. Rather, a complex combination of the two occurs. For all values of R in the region $R > R_{\min}$, however, the system does indeed lock onto the input frequency, thus displaying regional synchronous behavior.
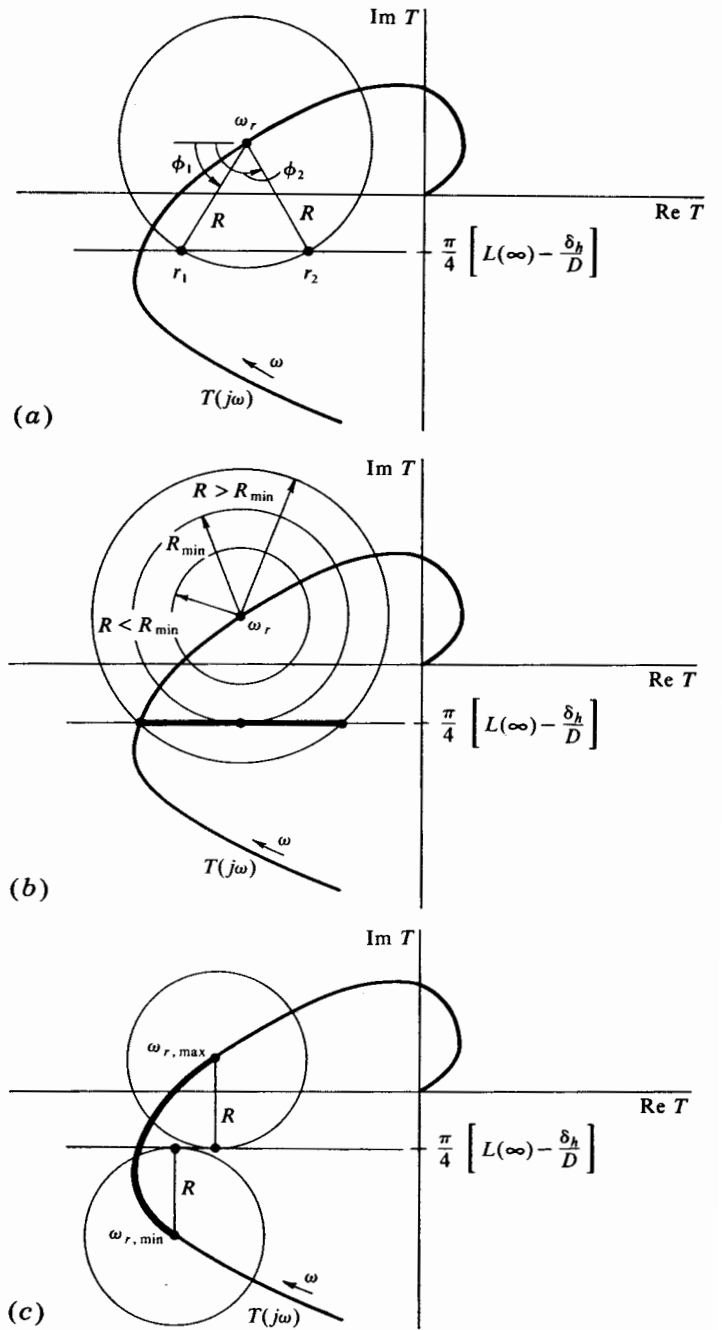
Figure 3.8-4 *Graphical solution of the forced response equations* (a). *Construction of region of synchronism for fixed input frequency* (b), *for fixed input amplitude* (c).

Viewed differently, for a fixed input amplitude R and varying input frequency ω_r, there will be some frequency region $\omega_{r,\max} > \omega_r > \omega_{r,\min}$ in which synchronous action can occur. Outside of this region, however, no such input-output synchronism is possible (Fig. 3.8-4c).

ANALYTICAL APPLICATION OF TSYPKIN'S METHOD

The problem we now face essentially reduces to the requirement for summing series of the form

and

$$\sum_{k=1,3,5,\dots}^{\infty} \mathrm{Re}\,[L_1(jk\omega_0)]$$

$$\sum_{k=1,3,5,\dots}^{\infty} \frac{1}{k}\,\mathrm{Im}\,[L_1(jk\omega_0)]$$

as encountered in Eq. (3.8-14), in terms of known functions. It happens that a summation procedure is available which solves this problem, having its basis in the theory of functions of a complex variable. By replacing a series of the form $\sum_{-\infty}^{\infty} f(n)$ with an appropriate contour integral, it is possible

TABLE 3.8-I   SUMMATION OF CERTAIN INFINITE SERIES

$L_1(s)$	$\displaystyle\sum_{k=1,3,5,\dots}^{\infty} \mathrm{Re}\,[L_1(jk\omega_0)]$	$\displaystyle\sum_{k=1,3,5,\dots}^{\infty} \frac{1}{k}\,\mathrm{Im}\,[L_1(jk\omega_0)]$
$\dfrac{1}{s+a}$	$\dfrac{\pi}{4\omega_0}\tanh\dfrac{\pi a}{2\omega_0}$	$-\dfrac{\pi}{4a}\tanh\dfrac{\pi a}{2\omega_0}$
$\dfrac{s+b}{(s+a)^2}$	$\dfrac{\pi}{8\omega_0\cosh^2(\pi a/2\omega_0)}$ $\times\left[\sinh\left(\dfrac{\pi a}{\omega_0}\right)-\dfrac{\pi}{\omega_0}(b-a)\right]$	$-\dfrac{\pi}{8a^2\cosh^2(\pi a/2\omega_0)}$ $\times\left[b\sinh\left(\dfrac{\pi a}{\omega_0}\right)-\dfrac{\pi a}{\omega_0}(b-a)\right]$
$\dfrac{(s+b_1)(s+b_2)}{(s+a)^3}$	$\dfrac{\pi}{16\omega_0\cosh^2(\pi a/2\omega_0)}$ $\times\left[2\sinh\left(\dfrac{\pi a}{\omega_0}\right)-\dfrac{\pi^2}{\omega_0{}^2}(a-b_1)\right.$ $\times(a-b_2)\tanh\left(\dfrac{\pi a}{2\omega_0}\right)$ $\left.+\dfrac{2\pi}{\omega_0}(2a-b_1-b_2)\right]$	$-\dfrac{\pi}{16a^3\cosh^2(\pi a/2\omega_0)}$ $\times\left[2b_1b_2\sinh\left(\dfrac{\pi a}{\omega_0}\right)-\pi^2(a-b_1)\right.$ $\times(a-b_2)\tanh\left(\dfrac{\pi a}{2\omega_0}\right)$ $\left.+2\pi\dfrac{a}{\omega_0}(a^2-b_1b_2)\right]$

Note: A useful relationship in applications of this table is

$$\frac{1}{(s+a)^2} = \lim_{b\to\infty}\left[\frac{1}{b}\frac{s+b}{(s+a)^2}\right]$$

to evaluate the series by application of the Cauchy residue formula. This manipulation, described elsewhere (Ref. 26), results in the formation of Table 3.8-1. In order to use this table, a given $L_1(s)$ is represented by its partial fraction expansion, and the terms are summed individually. Single-, double-, and triple-order poles can be accommodated by the three entries of the table.

Let us now consider the application of Tsypkin's method to a first-order relay control system which cannot be studied by the DF method (cf. Example 3.7-5).

Example 3.8-1 Investigate the possibility of limit cycle oscillations in the system of Fig. 3.8-5a. Consider time dimensioned in seconds.

For this system $L(\infty) = 0$; hence $L_1(s) = L(s)$. The first entry in Table 3.8-1 yields

$$\text{Re } [T(j\omega_0)] = \frac{2\pi}{\omega_0} \tanh \frac{2\pi}{\omega_0}$$

and

$$\text{Im } [T(j\omega_0)] = -\frac{\pi}{2} \tanh \frac{2\pi}{\omega_0}$$

From Eqs. (3.8-15) the conditions for sustained oscillation are

$$\frac{2\pi}{\omega_0} \tanh \frac{2\pi}{\omega_0} < \frac{2\pi}{\omega_0} \quad \text{or} \quad \tanh \frac{2\pi}{\omega_0} < 1$$

and

$$-\frac{\pi}{2} \tanh \frac{2\pi}{\omega_0} = -\frac{\pi}{5} \quad \text{or} \quad \tanh \frac{2\pi}{\omega_0} = 0.4$$

whence we observe that a limit cycle is indeed possible, and that its period is

$$T_0 = \frac{2\pi}{\omega_0} = \tanh^{-1} 0.4 = 0.424 \text{ sec}$$

The graphical construction is shown in Fig. 3.8-5b, from which we see that $d(\text{Im } T)/d\omega > 0$ at $\omega = \omega_0$. Hence the limit cycle is stable.

Tsypkin's method can be applied to the study of systems displaying time-delay (transport-lag) phenomena. In such applications we deal with factors of the form e^{-sT}, which do not appear in Table 3.8-1. A possible and frequently instructive approach with which to circumvent this problem is to replace the time-delay factors by suitable Padé approximants (Ref. 104), which are given as ratios of polynomials in s. Then Table 3.8-1 can be employed, and system oscillations can be studied. In the interest of retaining an exact version of Tsypkin's method wherever possible, we digress to discuss an alternative summation procedure.

Following Guillemin (Ref. 37), one can generate the sums of Table 3.8-2 in closed form by an elegant sequence of elementary Fourier series manipulations. We shall demonstrate here that this table can be simply extended to
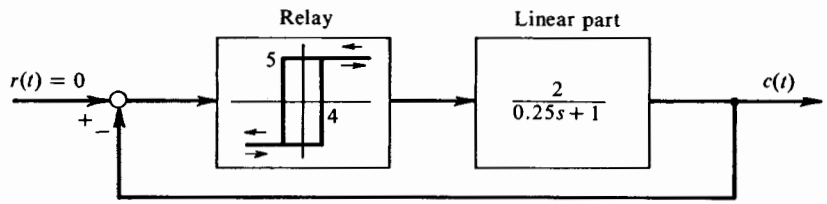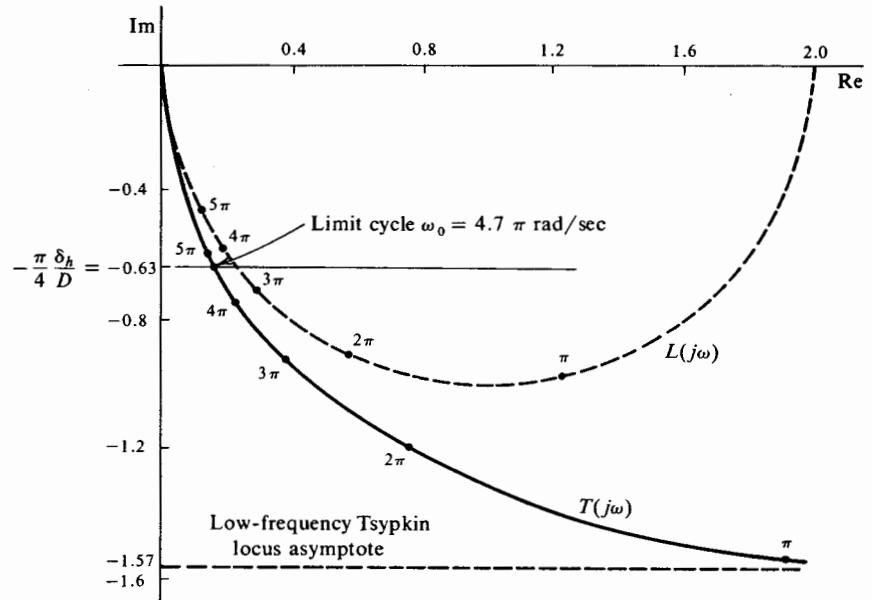
Figure 3.8-5 *(a) Limit cycling system.   (b) The graphical limit cycle solution.*

include all cases of interest not listed.   Observe that entry 3 can be derived from entry 2 merely by forming the function $f(x,a_1) - f(x,a_2)$ and performing the indicated algebra thereafter.   Similarly, it is an algebraic procedure to extend these results to any order denominator comprised of terms $(k^2 + a_i^2)$.   All series involving $\cos kx$ are derived from the related series involving $\sin kx$ by differentiation with respect to x.   A summed series containing one second-order-denominator root $(k^2 + a_i^2)^2$ can be obtained by differentiation with respect to a_i of the same series with that root to the first power.   Similarly, any-order-denominator root can be obtained by repeated differentiation.   Finally, note that any order numerator can be

TABLE 3.8-2   SUMMATION OF CERTAIN INFINITE SERIES $(0 < x < \pi)$

$$(1) \quad \sum_{k \text{ odd}}^{\infty} \frac{\sin kx}{k} = \frac{\pi}{4}$$

$$(2) \quad \sum_{k \text{ odd}}^{\infty} \frac{\sin kx}{k(k^2 + a^2)} = f(x,a)$$

$$(3) \quad \sum_{k \text{ odd}}^{\infty} \frac{\sin kx}{k(k^2 + a_1^2)(k^2 + a_2^2)} = \frac{f(x,a_1) - f(x,a_2)}{a_2^2 - a_1^2}$$

$$(4) \quad \sum_{k \text{ odd}}^{\infty} \frac{\cos kx}{k^2 + a^2} = f_x(x,a)$$

$$(5) \quad \sum_{k \text{ odd}}^{\infty} \frac{\cos kx}{(k^2 + a_1^2)(k^2 + a_2^2)} = \frac{f_x(x,a_1) - f_x(x,a_2)}{a_2^2 - a_1^2}$$

$$\text{where} \quad f(x,a) = \frac{\pi}{4} \frac{\cosh(\pi a/2) - \cosh(\pi a/2 - ax)}{a^2 \cosh(\pi a/2)}$$

$$f_x(x,a) = \frac{\partial f}{\partial x}(x,a) = \frac{\pi}{4} \frac{\sinh(\pi a/2 - ax)}{a \cosh(\pi a/2)}$$

developed by forming terms such as

$$bf(x,a_1) + cf(x,a_2) \qquad \text{or} \qquad bf_x(x,a_1) + cf_x(x,a_2)$$

from entries 2 and 4 of the table, respectively.

Example 3.8-2   Determine the amplitude and frequency of the limit cycle displayed by the system of Fig. 3.8-6.

For this system, since $L(\infty) = 0$, we have

$$L_1(s) = \frac{Ke^{-s\tau}}{s}$$

The Tsypkin locus is computed according to

$$\text{Re}\,[T(j\omega_0)] = -\frac{K}{\omega_0} \sum_{k=1,3,5,\ldots}^{\infty} \frac{\sin k\omega_0\tau}{k} = -\frac{K}{\omega_0}\frac{\pi}{4}$$

and

$$\text{Im}\,[T(j\omega_0)] = -\frac{K}{\omega_0} \sum_{k=1,3,5,\ldots}^{\infty} \frac{\cos k\omega_0\tau}{k^2} = -\frac{K}{\omega_0}\frac{\pi}{4}\left(\frac{\pi}{2} - \omega_0\tau\right)$$

where we have used entries 1 and 4 of Table 3.8-2, the latter in the limit as $a \to 0$. The conditions for sustained oscillation are therefore

$$-\frac{K}{\omega_0}\frac{\pi}{4} < 0$$

$$-\frac{K}{\omega_0}\frac{\pi}{4}\left(\frac{\pi}{2} - \omega_0\tau\right) = -\frac{\pi}{4}\frac{\delta_h}{D}$$
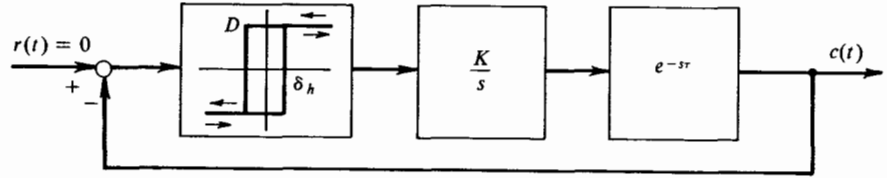
Figure 3.8-6   *Simple relay control system.*

the second of which yields, for $K > 0$,

$$T_0 = \frac{2\pi}{\omega_0} = 4\left(\tau + \frac{\delta_h}{DK}\right)$$

By physical argument the symmetrical limit cycle waveform is established as triangular with zero average value.   The peak amplitude A_0 is then the integrated nonlinearity output over one-quarter cycle.

$$A_0 = K\int_0^{T_0/4} D\,dt = \delta_h + DK\tau$$

METHODS OF HAMEL, BOHN, AND LUR'E

Additional methods for the exact solution for oscillations in relay control systems are now discussed briefly.

Hamel (Ref. 42) solves the problem of satisfying the required periodicity conditions by matching system trajectories in the phase plane.   The existence of a periodic oscillation is then defined by an intersection in the phase plane of the so-called Hamel locus (the locus of points at which a switch must occur in order to sustain an oscillation) with the switch line $x = -\delta_h$ (for a hysteretic relay).   It turns out that the Hamel and Tsypkin loci are fully equivalent in all systems with at least two more poles than zeros.   In fact, one can construct the Hamel locus directly by an abscissa scale change and a coordinate rotation of the Tsypkin locus.   Details of Hamel's method and the above-mentioned equivalence are available elsewhere (Ref. 30).

Bohn (Ref. 9) formulated the relay problem entirely in the time domain in order to overcome an apparent inadequacy in the Hamel and Tsypkin methods when applied to conservative systems.   Although Tsypkin (Ref. 107) has subsequently pointed out that this inadequacy does, in fact, not exist, Bohn's formulation is nevertheless of interest.   The result of this formulation is, for the hysteretic relay,

$$c(t)\bigg|_{t=0} = -\delta_h \quad \text{and} \quad \frac{dc}{dt}\bigg|_{t=T_0/2} > 0 \qquad (3.8\text{-}19)$$

where

$$c(t) = \frac{D}{\pi j}\oint_c \frac{L(s)}{s}\frac{\exp(st)}{1 + \exp(sT_0/2)}\,ds \qquad (3.8\text{-}20)$$

and the contour C encloses all left half-plane poles of $L(s)$, as well as the pole at the origin, but excludes all imaginary-axis pole pairs due to the zeros of $1 + \exp(sT_0/2)$. Proof of the validity of this formulation is left to the interested reader, who, upon closing the contour C in the right half-plane will also observe the re-creation of the usual Tsypkin conditions.

Lur'e (Ref. 63) treats the problem of steady oscillations directly in terms of the linear nth-order differential equation involved. By assuming a square-wave forcing of this differential equation, he formulates the nonlinear problem on a piecewise-linear basis. Then, by matching linear solutions at the switching instants, conditions for steady-state oscillations are determined. These conditions are presented in a compact canonic form in the reference cited, and are fully equivalent to the conditions derived by Bohn. *Aizerman* (Ref. 1) describes a method for the determination of self-oscillation modes in systems with an *arbitrary* piecewise-linear nonlinearity. This represents a degree of generalization beyond Lur'e's approach; however, in both instances the method of determining the periodic states reduces to a system of transcendental equations for which a general solution is not possible. These are in fact the same as the analytically summed Tsypkin conditions.

REFERENCES

1. Aizerman, M. A.: "Theory of Automatic Control," Pergamon Press, New York, 1963, chap. 5.
2. Andrievskii, V. R.: The Estimation of Self-oscillation Parameters in Nonlinear Automatic Control Systems, *Automation and Remote Control*, vol. 22 (September, 1961), pp. 146–150.
3. Bass, R. W.: Discussion of a paper by S. K. Chao, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 75 (September, 1956), p. 231.
4. Bass, R. W.: Equivalent Linearization, Nonlinear Circuit Synthesis and the Stabilization and Optimization of Control Systems, *Proc. Symp. Nonlinear Circuit Analysis*, MRI Series, Polytechnic Institute of Brooklyn, New York, 1956, pp. 163–198.
5. Bass, R. W.: Mathematical Legitimacy of Equivalent Linearization by Describing Functions, *Proc. IFAC*, Moscow, 1960, Butterworth Scientific Publications, London, pp. 2074–2083.
6. Bautin, N.: "Povedenie dinamicheskikh sistem vblizi granits ustojchivosti," Gostekhizdat, Moscow, 1949.
7. Bergen, A. R.: A Note on Tsypkin's Locus, *IRE Trans. Autom. Control*, vol. AC-7, no. 3 (April, 1962), pp. 78–80.
8. Bogoliubov, N. N., and Y. A. Mitropolsky: "Asymptotic Methods in the Theory of Nonlinear Oscillations" (translated from Russian), Gordon and Breach, Science Publishers, Inc., New York, 1961.
9. Bohn, E. V.: Stability Margins and Steady-state Oscillations in On-Off Feedback Systems, *IRE Trans. Circuit Theory*, vol. CT-8, no. 2 (June, 1961), pp. 127–130.
10. Bonenn, Ze'ev: Stability of Forced Oscillations in Nonlinear Feedback Systems, *IRE Trans. Autom. Control*, vol. AC-6 (December, 1958), pp. 109–111.
11. Bonenn, Ze'ev: Connection between Various Methods of Investigating Absolute Stability, *IRE Trans. Autom. Control*, vol. AC-5, no. 4 (September, 1960), pp. 330–331.

12. Bulgakov, B. V.: Periodic Processes in Free Pseudo-linear Oscillatory Systems, *J. Franklin Inst.* (Philadelphia, Pa.), vol. 235 (June, 1943), pp. 591–616.

13. Cahen, G.: Perturbations des oscillateurs filtres, *Comptes Rend. Acad. Sci.*, vol. 235 (December, 1952), pp. 1614–1617.

14. Casserly, G., and J. G. Truxal: Measurement and Stabilization of Nonlinear Feedback Systems, *IRE Conv. Record*, pt. 4, 1956, pp. 52–60.

15. Chao, S. K.: Design of a Contactor Servo Using Describing Function Theory, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 75 (September, 1956), pp. 223–233.

16. Chestnut, H.: Approximate Frequency-response Methods for Representing Saturation and Dead Band, in R. Oldenburger (ed.), "Frequency Response," The Macmillan Company, New York, 1956, pp. 269–287.

17. Choksy, N. H.: A New Analytic Stability Criterion for Use with Describing Functions, *JACC Proc.*, June, 1963, pp. 481–484.

18. Churchill, R. V.: "Complex Variables and Applications," 2d ed., McGraw-Hill Book Company, New York, 1960, chaps. 5 and 7.

19. Clauser, F. H.: A Theory of Nearly Linear Systems, *Proc. IFAC*, Moscow, Butterworth Scientific Publications, London, 1960, pp. 1557–1570.

20. Copson, E. T.: "Theory of Functions of a Complex Variable," Oxford University Press, New York, 1935.

21. Del Toro, V., and S. R. Parker: "Principles of Control System Engineering," McGraw-Hill Book Company, New York, 1960.

22. Dolgolenko, Yu. V.: The Approximate Derivation of Partial "Sliding" Periodic Modes in Relay Systems, *Automation and Remote Control*, vol. 18, no. 1 (April, 1958), pp. 1–27.

23. Embler, J. N., and C. H. Weaver: A New Method for Compensating Contactor Servomechanisms, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 76 (September, 1957), pp. 165–170.

24. Freeman, E. A.: The Effect of Speed-dependent Friction and Backlash on the Stability of Automatic Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 78 (January, 1959), pp. 680–691.

25. Fukuma, A., and M. Matsubara: Jump Resonance Criteria of Nonlinear Systems, *IEEE Trans. Autom. Control*, vol. AC-11, no. 4 (October, 1966), pp. 699–706.

26. Gelb, A., and L. J. Henrikson: On the Analytic Application of Tsypkin's Method, *IRE Trans. Autom. Control*, vol. AC-8 (January, 1963), pp. 69–70.

27. Gelb, A.: Limit Cycles in Symmetric Multiple Nonlinearity Systems, *IRE Trans. Autom. Control*, vol. AC-8, no. 2 (April, 1963), pp. 177–178.

28. Gibson, J. E., and F. B. Tuteur: The Response of Relay Amplifiers with Feedback, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 76 (November, 1957), pp. 303–307.

29. Gibson, J. E.: "Nonlinear Automatic Control," McGraw-Hill Book Company, New York, p. 426.

30. Gille, J. C., M. J. Pélegrin, and P. Decaulne: "Feedback Control Systems," McGraw-Hill Book Company, New York, 1959, chap. 26.

31. Glatenok, I. V.: Evaluation of the Region of the True Periodic Solution Found Approximately by the Harmonic Balance Method, *Automation and Remote Control*, vol. 18, no. 11 (September, 1958), pp. 1175–1178.

32. Goldfarb, L. C.: On Some Nonlinear Phenomena in Regulatory Systems, in R. Oldenburger (ed.), "Frequency Response," The Macmillan Company, New York, 1956, pp. 239–259.

33. Grabbe, E. M., S. Ramo, and E. E. Wooldridge: "Handbook of Automation, Computation and Control," John Wiley & Sons, Inc., vol. 1, chap. 25.

34. Graham, D., and D. McRuer: "Analysis of Nonlinear Control Systems," John Wiley & Sons, Inc., New York, 1961.

35. Greif, H. D.: Describing Function Method of Servomechanism Analysis Applied to Most Commonly Encountered Nonlinearities, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 72 (September, 1953), pp. 243–248.

36. Grensted, P. E. W.: The Frequency-response Analysis of Nonlinear Systems, *Proc. IEE*, vol. C-102 (April, 1955), pp. 244–253.

37. Guillemin, E. A.: "The Mathematics of Circuit Analysis," John Wiley & Sons, Inc., New York, 1958, pp. 479–482.

38. Gurski, R. J.: Bang-bang Control of an Acceleration Switching Hydraulic Servovalve, *1961 NEREM Record*, pp. 150–151.

39. Haas, V. B., Jr.: Coulomb Friction in Feedback Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 72 (1953), pp. 119–123.

40. Hale, J. K.: "Oscillations in Nonlinear Systems," McGraw-Hill Book Company, New York, 1963.

41. Halstenberg, R. V.: Combined Hysteresis and Nonlinear Gains in Complex Control Systems, *IRE Trans. Autom. Control*, vol. AC-3 (July, 1958), pp. 51–58.

42. Hamel, B.: Contribution a l'étude mathematique des systèmes de réglage par tout-ou-rien, *CEMV*, no. 17, Service Technique Aeronautique, 1949.

43. Hatanaka, H.: The Frequency Responses and Jump-resonance Phenomena of Nonlinear Feedback Control Systems, *Trans. ASME, J. Basic Eng.*, vol. D-85 (June, 1963), pp. 236–242.

44. Hill, J. C.: Closed-loop Response of Nonlinear Systems by a Functional Transformation Approach, *IRE Trans. Autom. Control*, vol. AC-7, no. 4 (July, 1962), pp. 39–44.

45. Holtzman, J. M.: Contraction Maps and Equivalent Linearization, *Bell System Tech. J.*, vol. 46, no. 10 (December, 1967), pp. 2405–2435.

46. Hopkin, A. M., and K. Ogata: An Analytic Frequency-response Solution for a Higher-order Servomechanism with a Nonlinear Control Element, *Trans. ASME, J. Basic Eng.*, vol. D-81 (March, 1959), pp. 41–45.

47. Jaffe, R. C.: Limit Cycles in Single and Multiple Loop On-Off Systems, *M.I.T. Instrumentation Lab. Rept.* T-358, 1963.

48. Johnson, E. C.: Sinusoidal Analysis of Feedback-control Systems Containing Nonlinear Elements, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 71 (July, 1952), pp. 169–181.

49. Kislyakov, V. S.: Application of the Krylov-Bogoliubov Method of Constructing Asymptotic Approximations to the Investigation of Systems with Lag, *Automation and Remote Control*, vol. 21 (November, 1960), pp. 307–315.

50. Kislyakov, V. S.: Foundations for the Application of the Harmonic Linearization Method to an Investigation of Periodic Oscillations in Systems with Lag, *Automation and Remote Control*, vol. 21 (May, 1961), pp. 1051–1056.

51. Korolev, N. A.: Pulse Stabilization of Automatic Control Relay Systems, *Automation and Remote Control*, vol. 18, no. 5 (April, 1958), pp. 435–446.

52. Korolev, N. A.: Approximate Determination of the Auto-oscillation Parameters in Relay Systems with Lagged Feedback, *Automation and Remote Control*, vol. 20, no. 11 (June, 1960), pp. 1430–1434.

53. Kochenburger, R. J.: Limiting in Feedback Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 72 (July, 1953), pp. 180–194.

54. Kochenburger, R. J.: A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, *Trans. AIEE*, pt. II, vol. 69 (1950), pp. 270–284.

55. Krasovskii, A. A.: On the Theory of Two-channel Follow-up Systems with a Relay Component in the AC Circuit, *Automation and Remote Control*, vol. 22 (April, 1961), pp. 911–919.

56. Landa, P. S., and S. P. Strelkov: On the Stability of an Aileron Control System Disturbed by Turbulence, *Automation and Remote Control*, vol. 22 (April, 1961), pp. 953–961.

57. Leonhard, A.: Extension of the Describing Function Method to the Investigation of Parametric Oscillations, *Proc. IFAC*, Basel, Switzerland, August, 1963, pp. 187/1–187/7.
58. Levinson, E.: Some Saturation Phenomena in Servomechanisms with Emphasis on the Tachometer Stabilized System, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 72 (March, 1953), pp. 1–9.
59. Levinson, E.: Gain-Phase Relations of Non-linear Circuits, *IRE Conv. Record*, pt. IV, 1958, pp. 141–159.
60. Levinson, E.: Non-linear Feedback Control Systems, Parts 3–5, *Electro-Technology*, vol. 72 (September, 1962), pp. 136–146; (October, 1962), pp. 131–140; (November, 1962), pp. 71–82.
61. Loeb, J. M.: Recent Advances in Non-linear Servo Theory, in R. Oldenburger (ed.), "Frequency Response," The Macmillan Company, New York, 1956, pp. 260–268.
62. Lozier, J. C.: A Steady-state Approach to the Theory of Saturable Servo Systems, *IRE Trans. Autom. Control*, vol. PGAC-1 (May, 1956), pp. 19–39.
63. Lur'e, A. I.: "Some Non-linear Problems in the Theory of Automatic Control," H.M. Stationery Office, London, 1957.
64. Magnus, K.: Approximate Criteria for Stability and Critical Point in Nonlinear Control Circuits, Metropolitan-Vickers translation 2257 of article in *Fachtagung Regelungstech., Contrib.* 71, Heidelberg, Germany, 1956, pp. 1–3.
65. Magnus, K.: On a Method of Investigating Nonlinear Systems of Oscillations and of Servomechanisms, NASA translation of original article in German (1955), OTS 59–16776, 1958.
66. Mahalanabis, A. K.: On Stabilization of Feedback Systems Affected by Hysteresis Nonlinearities, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 80 (November, 1961), pp. 277–285.
67. Manabe, S.: Analyze Nonlinear Controls with Saturation Functions, *Control Eng.*, vol. 8 (February, 1961), pp. 83–86.
68. McAllister, A. S.: A Graphical Method for Finding the Frequency Response of Nonlinear Closed-loop Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 80 (November, 1961), pp. 268–277.
69. Meyfarth, Philip F.: Bang-bang versus Linear Control of a Second-order Rate-type Servomotor, *Trans. ASME, J. Basic Eng.*, vol. D-82 (March, 1960), pp. 66–72.
70. Mikhail, S. L., and G. H. Fett: Frequency Response of Nonlinear Closed-loop Feedback Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 77 (November, 1958), pp. 436–438.
71. Minorsky, N.: "Nonlinear Oscillations," D. Van Nostrand Company, Inc., Princeton, N.J., 1962, chap. 14.
72. Mishkin, E., and L. Braun, Jr.: "Adaptive Control Systems," McGraw-Hill Book Company, New York, 1961, chap. 6.
73. Mishkin, E., and J. G. Truxal: Nonlinear Compensation Networks for Feedback Systems, *IRE Conv. Record*, pt. 4, 1957, pp. 3–7.
74. Mitropolsky, Y. A., and P. M. Senyk: Construction of an Asymptotic Solution for an Autonomic System with Strong Nonlinearity, OTS Publ. 62–15674, January, 1962.
75. Morosanov, I. S.: Methods of Extremum Control, *Automation and Remote Control*, vol. 18, no. 7 (June, 1958), pp. 1077–1092.
76. Morse, P. M., and H. Feshbach: "Methods of Theoretical Physics," McGraw-Hill Book Company, New York, 1953, pp. 413–414.
77. Moruzzi, R. L., and F. B. Tuteur: Nonlinear Servomechanism of Limited Dynamic Range, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 79 (November, 1960), pp. 314–320.
78. Mufti, I. H.: A Note on the Application of Harmonic Linearization, *IEEE Trans. Autom. Control*, vol. AC-8, no. 2 (April, 1963), pp. 175–177.

79. Naslin, P.: A Simplified Theory of Feedback Control Systems, Part II, *Process Control and Automation*, vol. 6 (May, 1959), pp. 206–209.

80. Naslin, P.: Summing Transmittances, *Process Control and Automation*, September, 1961, pp. 384–390.

81. Neimark, G.: O periodicheskikh rezhimakh i ustojchivosti relejnykh sistem, *Avtomatika i Telemekhanika*, vol. 14, no. 5 (1953), pp. 556–569.

82. Neimark, Yu. I., and L. P. Shil'nikov: On the Symmetric Periodic Motions of Multistage Relay Systems, *Automation and Remote Control*, vol. 20, no. 11 (June, 1960), pp. 1422–1429.

83. Nichols, N. B.: Backlash in a Velocity Lag Servomechanism, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 73 (January, 1954), pp. 462–467.

84. Numakura, T., T. Mivra, and A. Kamoi: Analyses of Servomechanisms by Describing Function Method on a Plane with Parameters as Coordinates, *AIEE Paper* CP-57-1035, 1957.

85. Ogata, K.: An Analytic Method for Finding the Closed-loop Frequency Response of Nonlinear Feedback Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 76 (November, 1957), pp. 277–285.

86. Pestel, E.: Anwendung der Ljapunovschen Methode und des Verfahrens von Krylov-Bogoliubov auf ein technisches Beispiel, in "Nichtliheare Regelungsvorgänge," R. Oldenbourg KG, Munich, 1956.

87. Petelin, D. P.: Approximate Determination of the Auto-oscillations in the Automatic Control System for a Synchronous Motor, *Automation and Remote Control*, vol. 20 (July, 1959), pp. 13–19.

88. Popov, E. P.: "The Dynamics of Automatic Control Systems," Addison-Wesley Publishing Company, Reading, Mass., 1962.

89. Popov, V. M.: Absolute Stability of Nonlinear Systems of Automatic Control, *Automation and Remote Control*, vol. 22 (February, 1961), pp. 857–875.

90. Popov, V. M., and A. Halaney: On the Stability of Nonlinear Automatic Control Systems with Lagging Argument, *Automation and Remote Control*, vol. 23 (February, 1963), pp. 783–786.

91. Sandberg, I. W.: On the Response of Nonlinear Control Systems to Periodic Input Signals, *Bell System Tech. J.*, vol. 43 (May, 1964), pp. 911–926.

92. Shulkind, D.: Accuracy Requirements for Nonlinear Compensation of Backlash, *IRE Trans. Autom. Control*, vol. AC-5, no. 2 (June, 1960), pp. 79–84.

93. Silberberg, M. Y., and P. M. Schultheiss: Stability Criteria for Some Nonlinear Servomechanisms, *AIEE Paper* DP 62–976, 1962.

94. Slemon, G. R.: A Method of Approximate Steady-state Analysis for Nonlinear Networks, *J. IEE*, London, pt. I, vol. 100 (September, 1953), pp. 275–287.

95. Smith, O. J. M.: "Feedback Control Systems," McGraw-Hill Book Company, New York, 1958.

96. Sparacino, R. R.: A Synthesis Technique for Quasi-optimal Relay Control, Sc.D. thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, Cambridge, Mass., September, 1961.

97. Starikova, M. V.: An Approximate Study of Self-oscillations in a Nonlinear Two-channel Automatic Control System, Wright-Patterson Air Force Base, Ohio, Foreign Technology Division, *Tech. Transl.* 62–833, June, 1962.

98. Stein, W. A., and G. J. Thaler: Obtaining the Frequency Response Characteristics of a Nonlinear Servomechanism from an Amplitude- and Frequency-sensitive Describing Function, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 77 (May, 1958), pp. 91–96.

99. Stoker, J. J.: "Non-linear Vibrations in Mechanical and Electrical Systems," Interscience Publishers, Inc., New York, 1950.

100. Syvi-Yan, T., and T. Lvi-Vy: Auto-oscillations in a Single-loop Automatic Control System Containing Two Symmetric Relays, *Automation and Remote Control*, vol. 20 (July, 1959), pp. 83–88.

101. Taft, V. A.: Stability of Periodic Conditions in Automatic Control Systems Found Approximately on the Basis of a Filter Hypothesis, *Automation and Remote Control*, vol. 19, no. 6 (March, 1959), pp. 550–555.

102. Thaler, G. J., and M. P. Pastel: "Analysis and Design of Nonlinear Feedback Control Systems," McGraw-Hill Book Company, New York, 1962.

103. Thomas, C. H.: Stability Characteristics of Closed-loop Systems with Dead-band, in R. Oldenburger (ed.), "Frequency Response," The Macmillan Company, New York, 1956, pp. 288–305.

104. Truxal, J. G.: "Automatic Feedback Control System Synthesis," McGraw-Hill Book Company, New York, 1955, pp. 548–550.

105. Tsypkin, Ya. Z.: Ob ustojchivosti periodicheskikh rezhimov v relejnykh systemakh avtomaticheskovo regulirovanija, *Avtomatika i Telemekhanika*, vol. 14, no. 5 (1953).

106. Tsypkin, Ya. Z.: "Teorija relejnykh sistem avtomaticheskovo regulirovanija," Gostekhizdat, Moscow, 1955, chaps. 8 and 9. Translation: "Sampling Systems Theory and Its Application," vols. 1 and 2, Pergamon Press and The Macmillan Company, New York, 1964.

107. Tsypkin, Ya. Z.: On the Determination of Steady-state Oscillations of On-Off Feedback Systems, *IRE Trans. Circuit Theory*, vol. CT-9, no. 3 (September, 1962).

108. Tustin, A.: The Effects of Backlash and of Speed-dependent Friction on the Stability of Closed-cycle Control Systems, *J. IEE*, London, pt. II, vol. 94 (May, 1947), pp. 143–151.

109. Urabe, M.: Galerkin's Procedure for Nonlinear Periodic Systems, *Arch. Rational Mech. Anal.*, vol. 20 (1965), pp. 120–152.

110. Van Horn, I. H.: Nonlinear Techniques Applied to the Analysis of Pilot Induced Oscillations, *IRE Conv. Record*, pt. 4, 1957, pp. 27–32.

111. Vavilov, A. A.: Investigation of Symmetric Self-oscillations in Relay Systems with Respect to Logarithmic Characteristics, *Automation and Remote Control*, vol. 24 (February, 1963), pp. 978–989.

112. West, J. C.: The Use of Frequency Response Analysis in Nonlinear Control Systems, in "Regelungstechnik, Moderne Theorien und ihre Verwendbarkeit," 1957 (in German).

113. West, J. C.: The Describing Function Technique Applied to Non-linear Systems, *Process Control and Automation*, vol. 7, pt. I (September, 1960), pp. 421–426; pt. II (October, 1960), pp. 469–473.

114. West, J. C., and P. N. Nikiforuk: The Frequency Response of a Servomechanism Designed for Optimum Transient Response, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 75 (September, 1956), pp. 234–239.

115. Willems, J.: "On the Stability of Systems Containing a Single Nonlinearity," M.S. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering, Cambridge, Mass., July, 1964.

PROBLEMS

3-1. A phase-shift oscillator has a cascaded *RC*-lag feedback network with transfer function

$$L(s) = \frac{K}{(\tau s + 1)^3}$$

The feedforward (amplifier) portion of the oscillator has a DF derived from consideration of both large and small oscillation amplitudes, given approximately by

$$N(A) \approx \begin{cases} \dfrac{5}{2+A} \ \underline{/-0.4A} \ \text{deg} & \text{for } A \leq 3 \\[2ex] \dfrac{3}{A} \ \underline{/-\dfrac{3.6}{A}} \ \text{deg} & \text{for } A > 3 \end{cases}$$

Plot the limit cycle amplitude and frequency as functions of τ (for $K = 4$) and as functions of K (for $\tau = 1$). What account need be taken of the amplifier phase contribution to the closed loop?

3-2. Using both polar and amplitude-phase plots, determine and study the stability of all limit cycles predicted by DF application to an ideal relay servo with linear elements given by

$$L(s) = \frac{Ke^{-sT}}{s}$$

Compute the exact amplitude and frequency of the single stable limit cycle, and thus show that the DF approximation leads to a perfect answer for both the waveform frequency and its fundamental amplitude component.

3-3. Investigate the existence and stability of limit cycle oscillations present in a servo with a limiter (abrupt saturation with drive levels $\pm D$ and input breakpoints $\pm\delta$) and linear elements

$$L(s) = \frac{K(\tau s + 1)^3}{s^3}$$

Use both analytical and graphical methods.

3-4. Investigate the possible appearance of steady-state oscillations in the equation

$$\ddot{\theta}(t) + a\dot{\theta}(t) + b\dot{\theta}(t - \tau) + \omega_0{}^2\theta(t) + g\theta^3(t) = 0$$

which arises in the study of a mechanism for minimizing the roll of a ship (W. J. Cunningham, "Introduction to Nonlinear Analysis," McGraw-Hill Book Company, New York, 1958, p. 333).

3-5. In a certain pneumatic control system the dominant system characteristics are known to be a nonlinear square-root-law downstream valve $[y(x) = K\sqrt{x}]$ and three cascaded isolated accumulators. Each of the accumulators is describable by a first-order lag of unity gain and time constant $\tau = 5$. What value of valve gain K do you recommend for a fast-responding but limit-cycle-free system?

3-6. In a closed-loop system, an amplifier, a motor, and a gear train are used in the forward loop, and a simple angular sensor mounted on the load is used to provide the feedback signal. The amplifier has an adjustable gain K. The motor has a $\dot{\theta}$/volt performance function $= 100/(1 + 2s)$ deg/sec/volt. The gear has a 10:1 reduction and a backlash of 5° at the input end. A pure viscous load is driven by the output shaft of the gear. The sensor has a sensitivity of 0.1 volt/deg. Study the limit cycle characteristic of this system as a function of K.

3-7. Find that value of gain K for which the closed-loop response of the system of Example 3.3-1 is down 3 db at a frequency of 10 radians/min. What closed-loop phase shift occurs at this frequency?

If the *nonlinear* integrator were replaced by an equivalent *linear* integrator, what would be the resultant closed-loop gain and phase shift at $\omega = 10$ radians/min?

3-8. Discuss the frequency response of the simple servomechanism of Fig. 3-1 when N is
  (a) A linear gain
  (b) A limiter (abrupt saturation)
  (c) A cubing nonlinearity
  Can jump resonance occur in any of the cases above? If not, devise a nonlinearity which will lead to jump resonance and discuss the resulting frequency response characteristics.
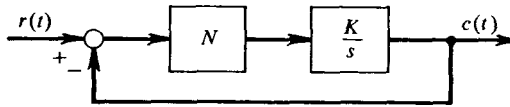

Figure 3-1 *Simple servomechanism for frequency response study.*

3-9. Sketch the closed-loop frequency response curves arising for the system of text Example 3.3-2 for the case in which the nonlinearity is a relay with drive levels $\pm D$ and dead zone 2δ. Consider the possibilities of both $M_r < \delta$ and $M_r > \delta$. What feedback compensation network can be employed to eliminate all jump resonances?

3-10. Demonstrate by manipulating the variables of Fig. 3-2 that the steady-state equation describing frequency response can be written as

$$\frac{M_r}{A} e^{-j\theta} - N(A) = L^{-1}(j\omega)$$

(*Hint:* Use of the notation $r = M_r e^{j\omega t}$, $x = A e^{j(\omega t + \theta)}$ may prove expedient.)
(Aizerman, *op. cit.*)

Describe a graphical procedure for determining A and θ as functions of ω, and illustrate a situation which depicts jump resonance possibilities.
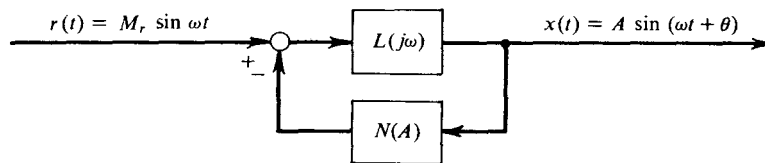
Figure 3-2 *Nonlinear system for general frequency response study.*

3-11. Using the critical frequency locus show that the system of Fig. 3-3 just fails to exhibit jump resonance phenomena if the following relationship holds:

$$\frac{\sqrt{ab}}{b - a} = \frac{1}{2\sqrt{3}}$$

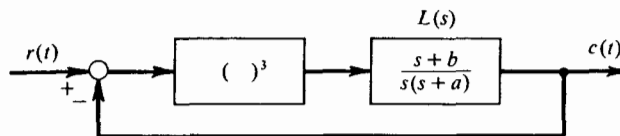[*Hint:* Work with the function $\varphi = \tan^{-1}(V/U)$, where $L = U + jV$.]

Figure 3-3   *System with a cubic nonlinearity.*

3-12. (*a*) Discuss the limit cycle behavior of the time-optimal control system whose frequency response was determined in Sec. 3.4.   Use the graphical approach.   On what physical basis can the DF result be justified?
(*b*) Discuss the graphical DF interpretation of conservative free oscillations of an ideal pendulum.

3-13. Design a compensation network with a single free parameter for a relay-controlled pure inertia plant $L(s) = K/s^2$, which provides a variable limit cycle frequency ω_0. What is the relationship between ω_0 and the free parameter?

3-14. A relay servomechanism motor-load transfer function is

$$\frac{\theta}{Y}(s) = \frac{100}{s(s+1)(s+10)} \qquad \text{radians/volt}$$

where θ is load position, and Y is the relay output (motor input).   Assuming a relay with hysteresis whose pull-in voltage is twice its drop-out voltage of 6 volts and whose drive levels are ± 45 volts, devise both lead and lag networks which ensure limit-cycle-free operation and discuss the resulting system in each case.   The linear feedback shaft encoder has a sensitivity of 1 volt/deg, and the compensation networks are to be placed in the feedforward path preceding the relay.

3-15. Study the limit cycle behavior of the system of Fig. 3-4.
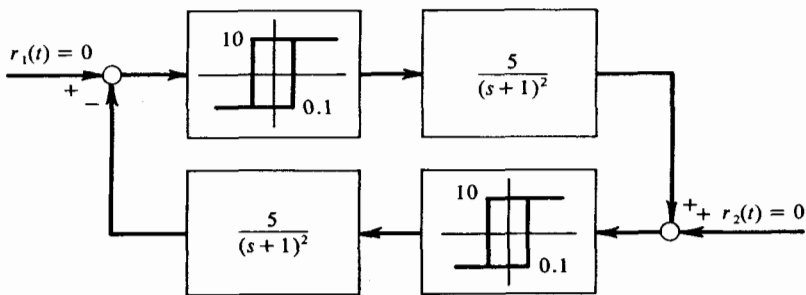
Figure 3-4   *Multiple nonlinearity system.*

3-16. By finding both exact and DF solutions for the self-oscillation frequency of the system of Fig. 3-5, demonstrate that the qualitative remarks regarding DF accuracy made in the text are again borne out.
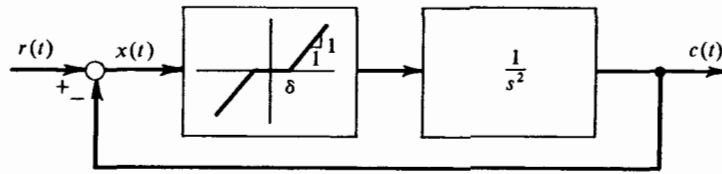
Figure 3-5 *Conservative second-order system with dead zone.*

3-17. For the following form of Rayleigh's equation,

$$\ddot{x} - \lambda(1 - \dot{x}^2)\dot{x} + x = 0$$

find the solution for limit cycle oscillations by the DF first approximation. By noting that the residual consists of a single term, construct the second-approximation DF solution, and thus show that the limit cycle frequency correct to second-order terms in λ is

$$\omega_0 \approx 1 + \frac{3\lambda^2 A^2}{64}$$

3-18. Using the DF method, compute first- and second-order approximation solutions for the limit cycle exhibited by the relay control system of Fig. 3-6.
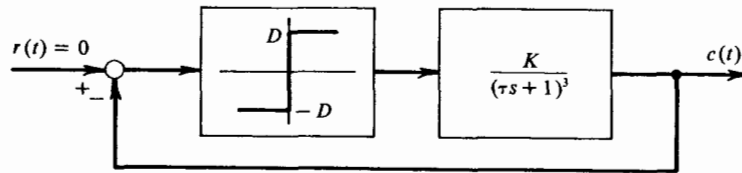
Figure 3-6 *Relay control system.*

3-19. Show that for all first- and second-order functions $L_1(s)$ the frequency calibration of the (coincident) linear elements' polar plot and corresponding Tsypkin locus in the high-frequency region tends toward a constant difference factor. [*Hint:* Consider the limit in increasing frequency of the function $|L_1(j\omega)/T(j\omega)|$.]

3-20. Discuss the nature of the limit cycle predicted by DF theory for the second-order ideal-relay system with $L(s) = K\omega_n^2/(s^2 + 2\zeta\omega_n s + \omega_n^2)$, and solve for the limit cycle exactly by use of Tsypkin's method. Repeat for $L(s) = K/[(s + 1)^3]$.

3-21. A satellite attitude control system with on-off thrust control and proportional plus rate feedback is shown in Fig. 3-7. Determine the limit cycle period for this system, and indicate probable accuracy of the result by tracing the relay-output third harmonic around the loop and comparing to the first-harmonic amplitude at the relay input.
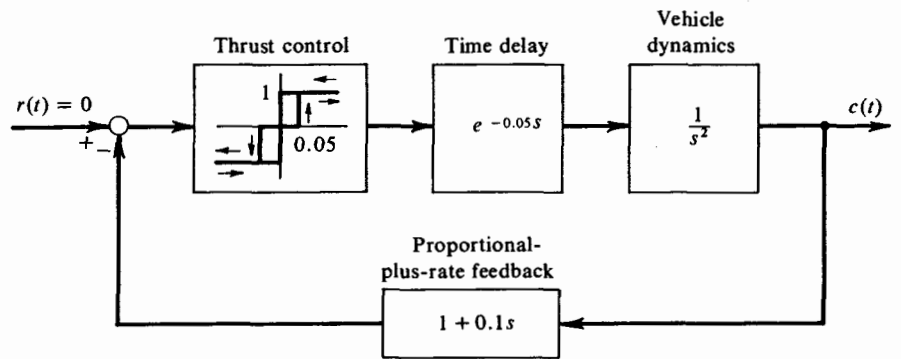
Figure 3-7   Satellite attitude control system.

3-22. Design the system of Fig. 3-8 to meet the specifications:
   (a) No limit cycle in the absence of input
   (b) Dead zone referred to the error $e(t) \leq 1$ unit
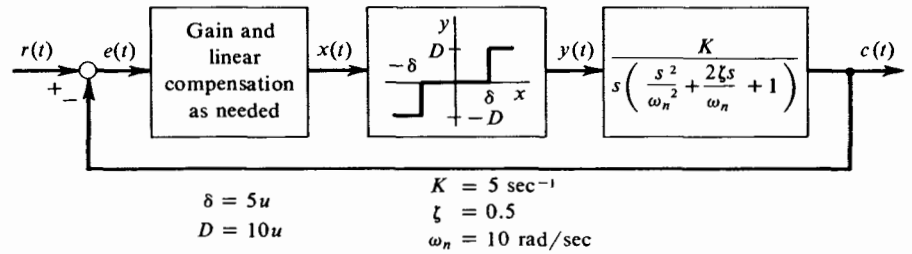

$$\delta = 5u$$
$$D = 10u$$

$$K = 5 \text{ sec}^{-1}$$
$$\zeta = 0.5$$
$$\omega_n = 10 \text{ rad/sec}$$

Figure 3-8

3-23. Show that the range of values of K for which the system of Fig. 3-9 will *not* limit-cycle is given by

$$0 \leq K < \frac{\pi \zeta \omega_n \delta}{D}$$

If K is twice the minimum value which results in a limit cycle, show that the resulting limit cycle amplitude and frequency are $A \approx 3.9\delta$, $\omega_0 = \omega_n$.
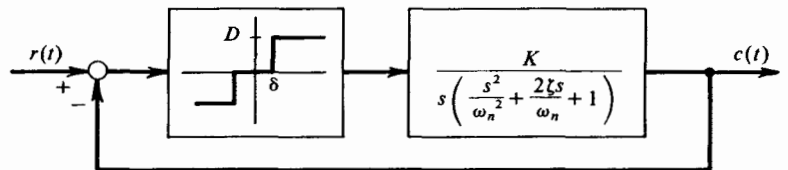
Figure 3-9

3-24. The relay-controlled system of Fig. 3-10 must meet the following specifications:
(a) It must have enough drive capability to follow input rates up to 50 units/sec.
(b) The limit cycle amplitude at $c(t)$ is to be no greater than 2 units.
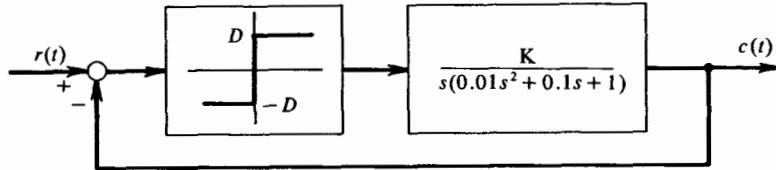Design appropriate compensation. You may place it either just before or just after the switch.

Figure 3-10

3-25. Sketch the input-output frequency response curves for amplitude and phase for the system of Fig. 3-11 with an input sinusoid of amplitude 15 units. You need not calculate the entire curve accurately. Just get some of the interesting points.
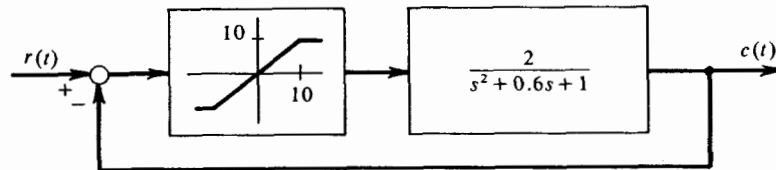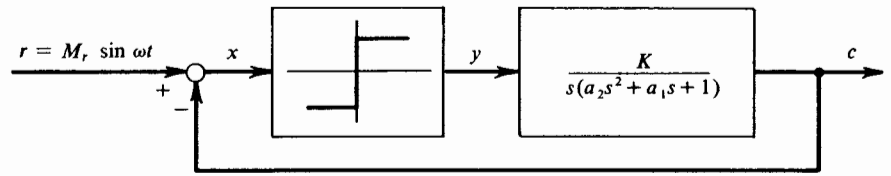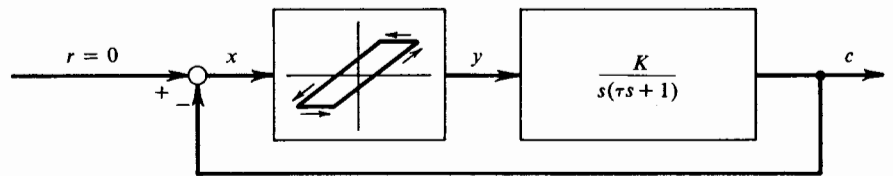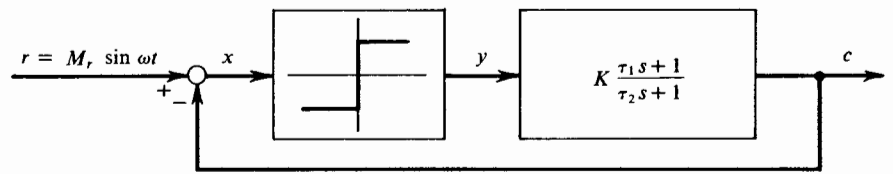

Figure 3-11

3-26. Would you expect single-sinusoid-input describing function theory to be useful in the applications of Fig. 3-12? If not, why not?
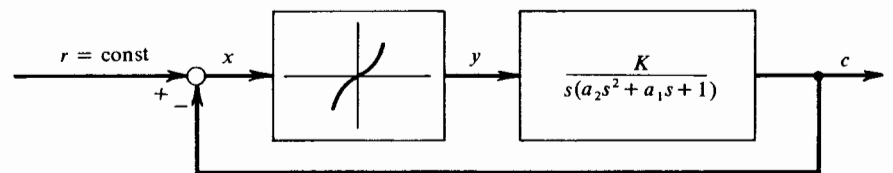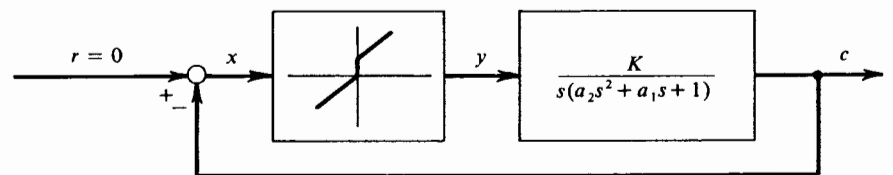
$$r = M_r \sin \omega t$$ x y $\dfrac{K}{s(a_2 s^2 + a_1 s + 1)}$ c

(a)

$$r = 0$$ x y $\dfrac{K}{s(\tau s + 1)}$ c

(b)

$$r = M_r \sin \omega t$$ x y $K\dfrac{\tau_1 s + 1}{\tau_2 s + 1}$ c

(c)

$$r = \text{const}$$ x y $\dfrac{K}{s(a_2 s^2 + a_1 s + 1)}$ c

(d)

$$r = 0$$ x y $\dfrac{K}{s(a_2 s^2 + a_1 s + 1)}$ c

(e)

Figure 3-12