Chapter 5
Describing Function Analysis

The frequency response method is a powerful tool for the analysis and design of linear control systems. It is based on describing a linear system by a complex-valued function, the frequency response, instead of a differential equation. The power of the method comes from a number of sources. First, *graphical* representations can be used to facilitate analysis and design. Second, *physical* insights can be used, because the frequency response functions have clear physical meanings. Finally, the method's complexity only increases mildly with system order. Frequency domain analysis, however, cannot be directly applied to nonlinear systems because frequency response functions cannot be defined for nonlinear systems.

Yet, for some nonlinear systems, an extended version of the frequency response method, called the *describing function method*, can be used to *approximately* analyze and predict nonlinear behavior. Even though it is only an approximation method, the desirable properties it inherits from the frequency response method, and the shortage of other systematic tools for nonlinear system analysis, make it an indispensable component of the bag of tools of practicing control engineers. The main use of describing function method is for the prediction of limit cycles in nonlinear systems, although the method has a number of other applications such as predicting subharmonics, jump phenomena, and the response of nonlinear systems to sinusoidal inputs.

This chapter presents an introduction to the describing function analysis of nonlinear systems. The basic ideas in the describing function method are presented in section 5.1. Section 5.2 discusses typical "hard nonlinearities" in control engineering, since describing functions are particularly useful for studying control systems containing such nonlinearities. Section 5.3 evaluates the describing functions for these hard nonlinearities. Section 5.4 is devoted to the description of how to use the describing function method for the prediction of limit cycles.

5.1 Describing Function Fundamentals

In this section, we start by presenting describing function analysis using a simple example, adapted from [Hsu and Meyer, 1968]. We then provide the formal definition of describing functions and some techniques for evaluating the describing functions of nonlinear elements.

5.1.1 An Example of Describing Function Analysis

The interesting and classical Van der Pol equation

$$\ddot{x} + \alpha(x^2 - 1)\dot{x} + x = 0 \tag{5.1}$$

(where α is a positive constant) has been treated by phase-plane analysis and Lyapunov analysis in the previous chapters. Let us now study it using a different technique, which shall lead us to the concept of a describing function. Specifically, let us determine whether there exists a limit cycle in this system and, if so, calculate the amplitude and frequency of the limit cycle (pretending that we have not seen the phase portrait of the Van der Pol equation in Chapter 2). To this effect, we first assume the existence of a limit cycle with undetermined amplitude and frequency, and then determine whether the system equation can indeed sustain such a solution. This is quite similar to the assumed-variable method in differential equation theory, where we first assume a solution of certain form, substitute it into the differential equation, and then attempt to determine the coefficients in the solution.

Before carrying out this procedure, let us represent the system dynamics in a block diagram form, as shown in Figure 5.1. It is seen that the feedback system in 5.1 contains a linear block and a nonlinear block, where the linear block, although unstable, has *low-pass* properties.

Now let us assume that there is a limit cycle in the system and the oscillation signal x is in the form of

$$x(t) = A\sin(\omega t)$$

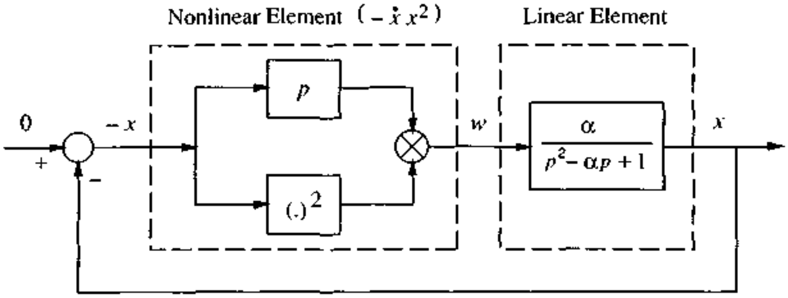

Figure 5.1 : Feedback interpretation of the Van der Pol oscillator

with A being the limit cycle amplitude and ω being the frequency. Thus,

$$\dot{x}(t) = A\,\omega\cos(\omega t)$$

Therefore, the output of the nonlinear block is

$$w = -x^2\,\dot{x} = -A^2\sin^2(\omega t)\;A\,\omega\cos(\omega t)$$

$$= -\frac{A^3\omega}{2}\,(\,1-\cos(2\,\omega t)\,)\,\cos(\omega t) = -\frac{A^3\omega}{4}\,(\cos(\omega t)-\cos(3\,\omega t)\,)$$

It is seen that w contains a third harmonic term. Since the linear block has low-pass properties, we can reasonably assume that this third harmonic term is sufficiently attenuated by the linear block and its effect is not present in the signal flow after the linear block. This means that we can approximate w by

$$w \approx -\frac{A^3}{4}\,\omega\cos\omega t = \frac{A^2}{4}\,\frac{\mathrm{d}}{\mathrm{d}t}\,[\,-A\sin(\omega t)\,]$$

so that the nonlinear block in Figure 5.1 can be approximated by the equivalent "quasi-linear" block in Figure 5.2. The "transfer function" of the quasi-linear block *depends on the signal amplitude* A, unlike a linear system transfer function (which is independent of the input magnitude).

In the frequency domain, this corresponds to

$$w = N(A,\omega)\,(\,-x\,) \tag{5.2}$$

where

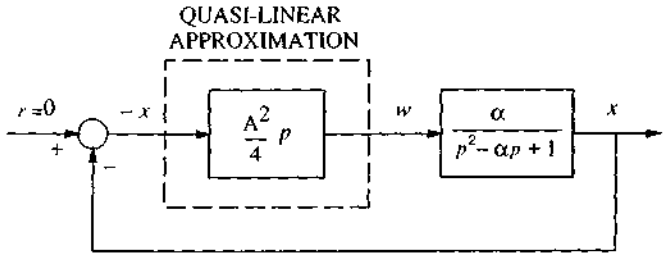

Figure 5.2 : Quasi-linear approximation of the Van der Pol oscillator

$$N(A, \omega) = \frac{A^2}{4}(j\omega)$$

That is, the nonlinear block can be approximated by the frequency response function $N(A, \omega)$. Since the system is assumed to contain a sinusoidal oscillation, we have

$$x = A\sin(\omega t) = G(j\omega)\,w = G(j\omega)\,N(A, \omega)\,(-x)$$

where $G(j\omega)$ is the linear component transfer function. This implies that

$$1 + \frac{A^2\,(j\omega)}{4}\,\frac{\alpha}{(j\omega)^2 - \alpha(j\omega) + 1} = 0$$

Solving this equation, we obtain

$$A = 2 \quad \omega = 1$$

Note that in terms of the Laplace variable p, the closed-loop characteristic equation of this system is

$$1 + \frac{A^2 p}{4}\,\frac{\alpha}{p^2 - \alpha p + 1} = 0 \tag{5.3}$$

whose eigenvalues are

$$\lambda_{1,2} = -\frac{1}{8}\alpha\,(A^2 - 4) \pm \sqrt{\frac{1}{64}\alpha^2\,(A^2 - 4)^2 - 1} \tag{5.4}$$

Corresponding to $A = 2$, we obtain the eigenvalues $\lambda_{1,2} = \pm j$. This indicates the existence of a limit cycle of amplitude 2 and frequency 1. It is interesting to note neither the amplitude nor the frequency obtained above depends on the parameter α in

Equation 5.1.

In the phase plane, the above approximate analysis suggests that the limit cycle is a circle of radius 2, regardless of the value of α. To verify the plausibility of this result, the real limit cycles corresponding to the different values of α are plotted (Figure 5.3). It is seen that the above approximation is reasonable for small value of α, but that the inaccuracy grows as α increases. This is understandable because as α grows the nonlinearity becomes more significant and the quasi-linear approximation becomes less accurate.

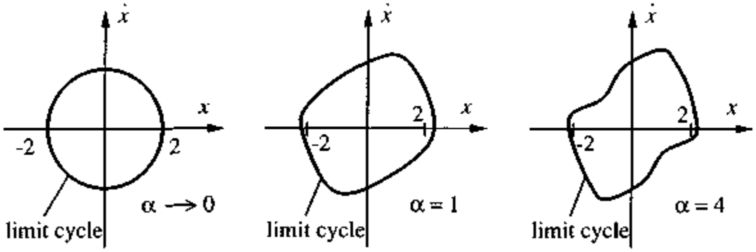

Figure 5.3 : Real limit cycles on the phase plane

The stability of the limit cycle can also be studied using the above analysis. Let us assume that the limit cycle's amplitude A is increased to a value larger than 2. Then, equation (5.4) shows that the closed-loop poles now have a negative real part. This indicates that the system becomes exponentially stable and thus the signal magnitude will decrease. Similar conclusions are obtained assuming that the limit cycle's amplitude A is decreased to a value less than 2. Thus, we conclude that the limit cycle is stable with an amplitude of 2.

Note that, in the above approximate analysis, the critical step is to replace the nonlinear block by the quasi-linear block which has the frequency response function $(A^2/4)\,(j\omega)$. Afterwards, the amplitude and frequency of the limit cycle can be determined from $1 + G(j\omega)\,N(A,\omega) = 0$. The function $N(A,\omega)$ is called the *describing function* of the nonlinear element. The above approximate analysis can be extended to predict limit cycles in other nonlinear systems which can be represented into the block diagram similar to Figure 5.1, as we shall do in section 5.4.

5.1.2 Applications Domain

Before moving on to the formal treatment of the describing function method, let us briefly discuss what kind of nonlinear systems it applies to, and what kind of information it can provide about nonlinear system behavior.

THE SYSTEMS

Simply speaking, any system which can be transformed into the configuration in Figure 5.4 can be studied using describing functions. There are at least two important classes of systems in this category.

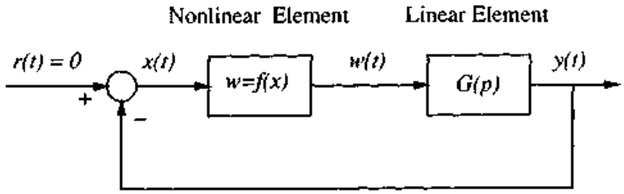

Figure 5.4 : A nonlinear system

The first important class consists of "almost" linear systems. By "almost" linear systems, we refer to systems which contain hard nonlinearities in the control loop but are otherwise linear. Such systems arise when a control system is designed using linear control but its implementation involves hard nonlinearities, such as motor saturation, actuator or sensor dead-zones, Coulomb friction, or hysteresis in the plant. An example is shown in Figure 5.5, which involves hard nonlinearities in the actuator.

Example 5.1: A system containing only one nonlinearity

Consider the control system shown in Figure 5.5. The plant is linear and the controller is also linear. However, the actuator involves a hard nonlinearity. This system can be rearranged into the form of Figure 5.4 by regarding $G_p G_1 G_2$ as the linear component G, and the actuator nonlinearity as the nonlinear element. □

"Almost" linear systems involving sensor or plant nonlinearities can be similarly rearranged into the form of Figure 5.4.

The second class of systems consists of genuinely nonlinear systems whose dynamic equations can actually be rearranged into the form of Figure 5.4. We saw an example of such systems in the previous section.

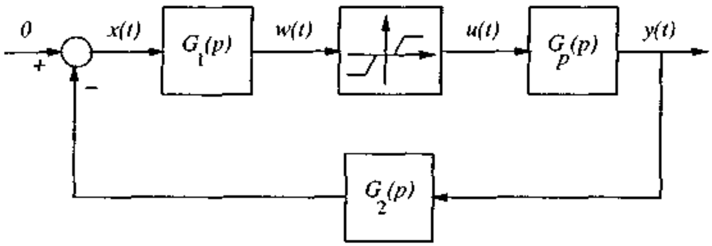

Figure 5.5 : A control system with hard nonlinearity

APPLICATIONS OF DESCRIBING FUNCTIONS

For systems such as the one in Figure 5.5, limit cycles can often occur due to the nonlinearity. However, linear control cannot predict such problems. Describing functions, on the other hand, can be conveniently used to discover the existence of limit cycles and determine their stability, regardless of whether the nonlinearity is "hard" or "soft." The applicability to limit cycle analysis is due to the fact that the form of the signals in a limit-cycling system is usually approximately *sinusoidal*. This can be conveniently explained on the system in Figure 5.4. Indeed, asssume that the linear element in Figure 5.4 has low-pass properties (which is the case of most physical systems). If there is a limit cycle in the system, then the system signals must all be periodic. Since, as a periodic signal, the input to the linear element in Figure 5.4 can be expanded as the sum of many harmonics, and since the linear element, because of its low-pass property, filters out higher frequency signals, the output $y(t)$ must be composed mostly of the lowest harmonics. Therefore, it is appropriate to assume that the signals in the whole system are basically sinusoidal in form, thus allowing the technique in subsection 5.1.1 to be applied.

Prediction of limit cycles is very important, because limit cycles can occur frequently in physical nonlinear system. Sometimes, a limit cycle can be desirable. This is the case of limit cycles in the electronic oscillators used in laboratories. Another example is the so-called dither technique which can be used to minimize the negative effects of Coulomb friction in mechanical systems. In most control systems, however, limit cycles are undesirable. This may be due to a number of reasons:

1. limit cycle, as a way of instability, tends to cause poor control accuracy
2. the constant oscillation associated with the limit cycles can cause increasing wear or even mechanical failure of the control system hardware
3. limit cycling may also cause other undesirable effects, such as passenger

discomfort in an aircraft under autopilot

In general, although a precise knowledge of the waveform of a limit cycle is usually not mandatory, the knowledge of the limit cycle's existence, as well as that of its approximate amplitude and frequency, is critical. The describing function method can be used for this purpose. It can also guide the design of compensators so as to avoid limit cycles.

5.1.3 Basic Assumptions

Consider a nonlinear system in the general form of Figure 5.4. In order to develop the *basic version* of the describing function method, the system has to satisfy the following four conditions:

1. *there is only a single nonlinear component*
2. *the nonlinear component is time-invariant*
3. *corresponding to a sinusoidal input* $x = \sin(\omega t)$ *, only the fundamental component* $w_1(t)$ *in the output* $w(t)$ *has to be considered*
4. *the nonlinearity is odd*

The first assumption implies that if there are two or more nonlinear components in a system, one either has to lump them together as a single nonlinearity (as can be done with two nonlinearities in parallel), or retain only the primary nonlinearity and neglect the others.

The second assumption implies that we consider only autonomous nonlinear systems. It is satisfied by many nonlinearities in practice, such as saturation in amplifiers, backlash in gears, Coulomb friction between surfaces, and hysteresis in relays. The reason for this assumption is that the Nyquist criterion, on which the describing function method is largely based, applies only to linear time-invariant systems.

The third assumption is the *fundamental assumption* of the describing function method. It represents an *approximation*, because the output of a nonlinear element corresponding to a sinusoidal input usually contains higher harmonics besides the fundamental. This assumption implies that the higher-frequency harmonics can all be neglected in the analysis, as compared with the fundamental component. For this assumption to be valid, it is important for the linear element following the nonlinearity to have low-pass properties, *i.e.*,

$$|G(j\omega)| >> |G(jn\omega)| \quad \text{for} \quad n = 2, 3, ... \tag{5.5}$$

This implies that higher harmonics in the output will be filtered out significantly. Thus, the third assumption is often referred to as the *filtering hypothesis*.

The fourth assumption means that the plot of the nonlinearity relation $f(x)$ between the input and output of the nonlinear element is symmetric about the origin. This assumption is introduced for simplicity, *i.e.*, so that the static term in the Fourier expansion of the output can be neglected. Note that the common nonlinearities discussed before all satisfy this assumption.

The relaxation of the above assumptions has been widely studied in literature, leading to describing function approaches for general situations, such as multiple nonlinearities, time-varying nonlinearities, or multiple-sinusoids. However, these methods based on relaxed conditions are usually much more complicated than the basic version, which corresponds to the above four assumptions. In this chapter, we shall mostly concentrate on the basic version.

5.1.4 Basic Definitions

Let us now discuss how to represent a *nonlinear component* by a describing function. Let us consider a sinusoidal input to the nonlinear element, of amplitude A and frequency ω, *i.e.*, $x(t) = A\sin(\omega t)$, as shown in Figure 5.6. The output of the nonlinear component $w(t)$ is often a periodic, though generally non-sinusoidal, function. Note that this is always the case if the nonlinearity $f(x)$ is single-valued, because the output is $f[A\sin(\omega(t+2\pi/\omega))] = f[A\sin(\omega t)]$. Using Fourier series, the periodic function $w(t)$ can be expanded as

$$w(t) = \frac{a_o}{2} + \sum_{n=1}^{\infty} [a_n \cos(n\omega t) + b_n \sin(n\omega t)] \tag{5.6}$$

where the Fourier coefficients a_i's and b_i's are generally functions of A and ω, determined by

$$a_o = \frac{1}{\pi}\int_{-\pi}^{\pi} w(t)\, d(\omega t) \tag{5.7a}$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} w(t)\cos(n\omega t)\, d(\omega t) \tag{5.7b}$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} w(t)\sin(n\omega t)\, d(\omega t) \tag{5.7c}$$

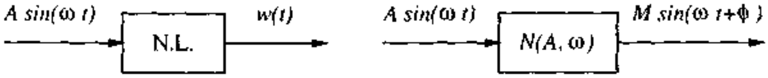

Figure 5.6 : A nonlinear element and its describing function representation

Due to the fourth assumption above, one has $a_o = 0$. Furthermore, the third assumption implies that we only need to consider the fundamental component $w_1(t)$, namely

$$w(t) \approx w_1(t) = a_1 \cos(\omega t) + b_1 \sin(\omega t) = M \sin(\omega t + \phi) \tag{5.8}$$

where

$$M(A,\omega) = \sqrt{a_1{}^2 + b_1{}^2} \quad \text{and} \quad \phi(A,\omega) = \arctan(a_1/b_1).$$

Expression (5.8) indicates that the fundamental component corresponding to a sinusoidal input is a sinusoid at the same frequency. In complex representation, this sinusoid can be written as $w_1 = M e^{j(\omega t + \phi)} = (b_1 + j a_1)\, e^{j\omega t}$.

Similarly to the concept of frequency response function, which is the frequency-domain ratio of the sinusoidal input and the sinusoidal output of a system, we define the *describing function* of the nonlinear element to be *the complex ratio of the fundamental component of the nonlinear element by the input sinusoid, i.e.,*

$$N(A,\omega) = \frac{M e^{j(\omega t + \phi)}}{A e^{j\omega t}} = \frac{M}{A} e^{j\phi} = \frac{1}{A}(b_1 + j a_1) \tag{5.9}$$

With a describing function representing the nonlinear component, the nonlinear element, in the presence of sinusoidal input, can be treated as if it were a linear element with a frequency response function $N(A,\omega)$, as shown in Figure 5.6. The concept of a describing function can thus be regarded as an extension of the notion of frequency response. For a linear dynamic system with frequency response function $H(j\omega)$, the describing function is independent of the input gain, as can be easily shown. However, the describing function of a nonlinear element differs from the frequency response function of a linear element in that it depends on the input amplitude A. Therefore, representing the nonlinear element as in Figure 5.6 is also called quasi-linearization.

Generally, the describing function depends on the frequency and amplitude of the input signal. There are, however, a number of special cases. When the nonlinearity is *single-valued*, the describing function $N(A,\omega)$ is *real and independent of the input frequency* ω. The realness of N is due to the fact that $a_1 = 0$, which is true because $f[A\sin(\omega t)]\cos(\omega t)$, the integrand in the expression (5.7b) for a_1, is an odd function of ωt, and the domain of integration is the symmetric interval $[-\pi, \pi]$. The frequency-independent nature is due to the fact that the integration of the single-valued function $f[A\sin(\omega t)]\sin(\omega t)$ in expression (5.7c) is done for the variable ωt, which implies that ω does not explicitly appear in the integration.

Although we have implicitly assumed the nonlinear element to be a scalar nonlinear function, the definition of the describing function also applies to the case when the nonlinear element contains dynamics (*i.e.*, is described by differential equations instead of a function). The derivation of describing functions for such nonlinear elements is usually more complicated and may require experimental evaluation.

5.1.5 Computing Describing Functions

A number of methods are available to determine the describing functions of nonlinear elements in control systems, based on definition (5.9). We now briefly describe three such methods: analytical calculation, experimental determination, and numerical integration. Convenience and cost in each particular application determine which method should be used. One thing to remember is that precision is not critical in evaluating describing functions of nonlinear elements, because the describing function method is itself an approximate method.

ANALYTICAL CALCULATION

When the nonlinear characteristics $w = f(x)$ (where x is the input and w the output) of the nonlinear element are described by an explicit function and the integration in (5.7) can be easily carried out, then analytical evaluation of the describing function based on (5.7) is desirable. The explicit function $f(x)$ of the nonlinear element may be an idealized representation of simple nonlinearities such as saturation and dead-zone, or it may be the curve-fit of an input-output relationship for the element. However, for nonlinear elements which evade convenient analytical expressions or contain dynamics, the analytical technique is difficult.

NUMERICAL INTEGRATION

For nonlinearities whose input-output relationship $w = f(x)$ is given by graphs or tables, it is convenient to use numerical integration to evaluate the describing functions. The idea is, of course, to approximate integrals in (5.7) by discrete sums over small intervals. Various numerical integration schemes can be applied for this purpose. It is obviously important that the numerical integration be easily implementable by computer programs. The result is a plot representing the describing function, which can be used to predict limit cycles based on the method to be developed in section 5.4.

EXPERIMENTAL EVALUATION

The experimental method is particularly suitable for complex nonlinearities and dynamic nonlinearities. When a system nonlinearity can be isolated and excited with sinusoidal inputs of known amplitude and frequency, experimental determination of the describing function can be obtained by using a harmonic analyzer on the output of the nonlinear element. This is quite similar to the experimental determination of frequency response functions for linear elements. The difference here is that not only the frequencies, but also the *amplitudes* of the input sinusoidal should be varied. The results of the experiments are a set of curves on complex planes representing the describing function $N(A, \omega)$, instead of analytical expressions. Specialized instruments are available which automatically compute the describing functions of nonlinear elements based on the measurement of nonlinear element response to harmonic excitation.

Let us illustrate on a simple nonlinearity how to evaluate describing functions using the analytical technique.

Example 5.2: Describing function of a hardening spring

The characteristics of a hardening spring are given by

$$w = x + x^3/2$$

with x being the input and w being the output. Given an input $x(t) = A\sin(\omega t)$, the output $w(t) = A\sin(\omega t) + A^3\sin^3(\omega t)/2$ can be expanded as a Fourier series, with the fundamental being

$$w_1(t) = a_1\cos\omega t + b_1\sin\omega t$$

Because $w(t)$ is an odd function, one has $a_1 = 0$, according to (5.7). The coefficient b_1 is

$$b_1 = \frac{1}{\pi}\int_{-\pi}^{\pi} [A\sin(\omega t) + A^3\sin^3(\omega t)/2]\sin(\omega t)\, d(\omega t) = A + \frac{3}{8}A^3$$

Therefore, the fundamental is

$$w_1 = (A + \frac{3}{8}A^3)\sin(\omega t)$$

and the describing function of this nonlinear component is

$$N(A,\omega) = N(A) = 1 + \frac{3}{8}A^2$$

Note that due to the odd nature of this nonlinearity, the describing function is real, being a function only of the amplitude of the sinusoidal input. □

5.2 Common Nonlinearities In Control Systems

In this section, we take a closer look at the nonlinearities found in control systems. Consider the typical system block shown in Figure 5.7. It is composed of four parts: a plant to be controlled, sensors for measurement, actuators for control action, and a control law, usually implemented on a computer. Nonlinearities may occur in any part of the system, and thus make it a nonlinear control system.

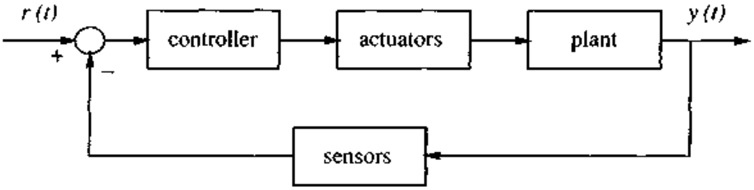

Figure 5.7 : Block diagram of control systems

CONTINUOUS AND DISCONTINUOUS NONLINEARITIES

Nonlinearities can be classified as *continuous* and *discontinuous*. Because discontinuous nonlinearities cannot be locally approximated by linear functions, they are also called "hard" nonlinearities. Hard nonlinearities are commonly found in control systems, both in small range operation and large range operation. Whether a system in small range operation should be regarded as nonlinear or linear depends on the magnitude of the hard nonlinearities and on the extent of their effects on the system performance.

Because of the common occurence of hard nonlinearities, let us briefly discuss the characteristics and effects of some important ones.

Saturation

When one increases the input to a physical device, the following phenomenon is often observed: when the input is small, its increase leads to a corresponding (often proportional) increase of output; but when the input reaches a certain level, its further increase does produces little or no increase of the output. The output simply stays around its maximum value. The device is said to be in *saturation* when this happens. Simple examples are transistor amplifiers and magnetic amplifiers. A saturation nonlinearity is usually caused by limits on component size, properties of materials, and available power. A typical saturation nonlinearity is represented in Figure 5.8, where the thick line is the real nonlinearity and the thin line is an idealized saturation nonlinearity.

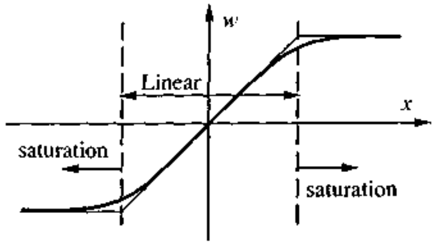

Figure 5.8 : A saturation nonlinearity

Most actuators display saturation characteristics. For example, the output torque of a two-phase servo motor cannot increase infinitely and tends to saturate, due to the properties of the magnetic material. Similarly, valve-controlled hydraulic servo motors are saturated by the maximum flow rate.

Saturation can have complicated effects on control system performance. Roughly speaking, the occurence of saturation amounts to reducing the gain of the device (*e.g.*, the amplifier) as the input signals are increased. As a result, if a system is unstable in its linear range, its divergent behavior may be suppressed into a self-sustained oscillation, due to the inhibition created by the saturating component on the system signals. On the other hand, in a linearly stable system, saturation tends to slow down the response of the system, because it reduces the effective gain.

On-off nonlinearity

An extreme case of saturation is the *on-off* or relay nonlinearity. It occurs when the linearity range is shrunken to zero and the slope in the linearity range becomes vertical. Important examples of on-off nonlinearities include output torques of gas jets for spacecraft control (as in example 2.5) and, of course, electrical relays. On-off nonlinearities have effects similar to those of saturation nonlinearities. Furthermore they can lead to "chattering" in physical systems due to their discontinuous nature.

Dead-zone

In many physical devices, the output is zero until the magnitude of the input exceeds a certain value. Such an input-output relation is called a *dead-zone*. Consider for instance a d.c. motor. In an idealistic model, we assume that any voltage applied to the armature windings will cause the armature to rotate, with small voltage causing small motion. In reality, due to the static friction at the motor shaft, rotation will occur only if the torque provided by the motor is sufficiently large. Similarly, when transmitting motion by connected mechanical components, dead zones result from manufacturing clearances. Similar dead-zone phenomena occur in valve-controlled pneumatic actuators and in hydraulic components.

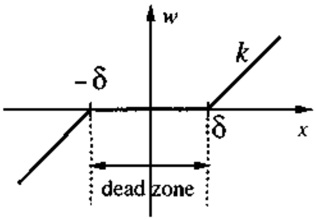

Figure 5.9 : A dead-zone nonlinearity

Dead-zones can have a number of possible effects on control systems. Their most common effect is to decrease static output accuracy. They may also lead to limit cycles or system instability because of the lack of response in the dead zone. In some cases, however, they may actually stabilize a system or suppress self-oscillations. For example, if a dead-zone is incorporated into an ideal relay, it may lead to the avoidance of the oscillation at the contact point of the relay, thus eliminating sparks and reducing wear at the contact point. In chapter 8, we describe a dead-zone technique to improve the robustness of adaptive control systems with respect to measurement noise.

Backlash and hysteresis

Backlash often occurs in transmission systems. It is caused by the small gaps which exist in transmission mechanisms. In gear trains, there always exist small gaps between a pair of mating gears, due to the unavoidable errors in manufacturing and assembly. Figure 5.10 illustrates a typical situation. As a result of the gaps, when the driving gear rotates a smaller angle than the gap b, the driven gear does not move at all, which corresponds to the dead-zone (OA segment in Figure 5.10); after contact has been established between the two gears, the driven gear follows the rotation of the driving gear in a linear fashion (AB segment). When the driving gear rotates in the reverse direction by a distance of $2b$, the driven gear again does not move, corresponding to the BC segment in Figure 5.10. After the contact between the two gears is re-established, the driven gear follows the rotation of the driving gear in the reverse direction (CD segment). Therefore, if the driving gear is in periodic motion, the driven gear will move in the fashion represented by the closed path EBCD. Note that the height of B, C, D, E in this figure depends on the amplitude of the input sinusoidal.

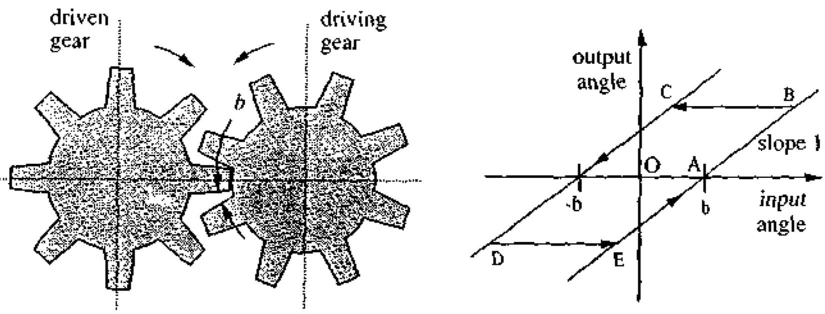

Figure 5.10 : A backlash nonlinearity

A critical feature of backlash is its multi-valued nature. Corresponding to each input, two output values are possible. Which one of the two occur depends on the history of the input. We remark that a similar multi-valued nonlinearity is hysteresis, which is frequently observed in relay components.

Multi-valued nonlinearities like backlash and hysteresis usually lead to energy storage in the system. Energy storage is a frequent cause of instability and self-sustained oscillation.

5.3 Describing Functions of Common Nonlinearities

In this section, we shall compute the describing functions for a few common nonlinearities. This will not only allow us to familiarize ourselves with the frequency domain properties of these common nonlinearities, but also will provide further examples of how to derive describing functions for nonlinear elements.

SATURATION

The input-output relationship for a saturation nonlinearity is plotted in Figure 5.11, with a and k denoting the range and slope of the linearity. Since this nonlinearity is single-valued, we expect the describing function to be a real function of the input

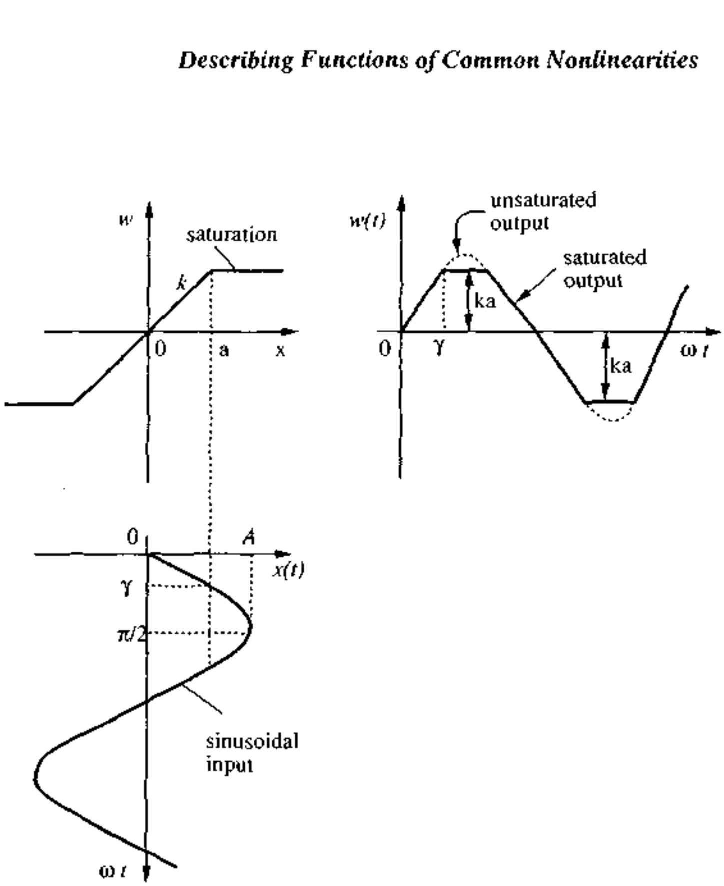

Figure 5.11 : Saturation nonlinearity and the corresponding input-output relationship

amplitude.

Consider the input $x(t) = A\sin(\omega t)$. If $A \leq a$, then the input remains in the linear range, and therefore, the output is $w(t) = kA\sin(\omega t)$. Hence, the describing function is simply a constant k.

Now consider the case $A > a$. The input and the output functions are plotted in Figure 5.11. The output is seen to be symmetric over the four quarters of a period. In the first quarter, it can be expressed as

$$w(t) = \begin{cases} kA\sin(\omega t) & 0 \leq \omega t \leq \gamma \\ ka & \gamma < \omega t \leq \pi/2 \end{cases}$$

where $\gamma = \sin^{-1}(a/A)$. The odd nature of $w(t)$ implies that $a_1 = 0$ and the symmetry over the four quarters of a period implies that

$$b_1 = \frac{4}{\pi}\int_o^{\pi/2} w(t)\sin(\omega t)\, d(\omega t)$$

$$= \frac{4}{\pi}\int_o^{\gamma} kA\sin^2(\omega t)\, d(\omega t) + \frac{4}{\pi}\int_{\gamma}^{\pi/2} ka\sin(\omega t)\, d(\omega t)$$

$$= \frac{2kA}{\pi}\,[\,\gamma + \frac{a}{A}\sqrt{1-\frac{a^2}{A^2}}\,] \tag{5.10}$$

Therefore, the describing function is

$$N(A) = \frac{b_1}{A} = \frac{2k}{\pi}\,[\sin^{-1}\frac{a}{A} + \frac{a}{A}\sqrt{1-\frac{a^2}{A^2}}\,] \tag{5.11}$$

The normalized describing function ($N(A)/k$) is plotted in Figure 5.12 as a function of A/a. One can observe three features for this describing function:

1. $N(A) = k$ if the input amplitude is in the linearity range
2. $N(A)$ decreases as the input amplitude increases
3. there is no phase shift

The first feature is obvious, because for small signals the saturation is not displayed. The second is intuitively reasonable, since saturation amounts to reduce the ratio of the output to input. The third is also understandable because saturation does not cause the delay of the response to input.

As a special case, one can obtain the describing function for the relay-type (on-off) nonlinearity shown in Figure 5.13. This case corresponds to shrinking the linearity range in the saturation function to zero, *i.e.*, $a \to 0$, $k \to \infty$, but $ka = M$. Though b_1 can be obtained from (5.10) by taking the limit, it is more easily obtained directly as

$$b_1 = \frac{4}{\pi}\int_0^{\pi/2} M\sin(\omega t)\, d(\omega t) = \frac{4}{\pi}M$$

Therefore, the describing function of the relay nonlinearity is

$$N(A) = \frac{4M}{\pi A} \tag{5.12}$$

The normalized describing function (N/M) is plotted in Figure 5.13 as a function of

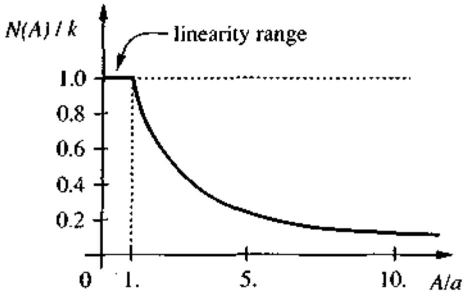

Figure 5.12 : Describing function of the saturation nonlinearity

input amplitude. Although the describing function again has no phase shift, the flat segment seen in Figure 5.12 is missing in this plot, due to the completely nonlinear nature of the relay. The asymptic properties of the describing function curve in Figure 5.13 are particularly interesting. When the input is infinitely small, the describing function is infinitely large. When the input is infinitely large, the describing function is infinitely small. One can gain an intuitive understanding of these properties by considering the ratio of the output to input for the on-off nonlinearity.

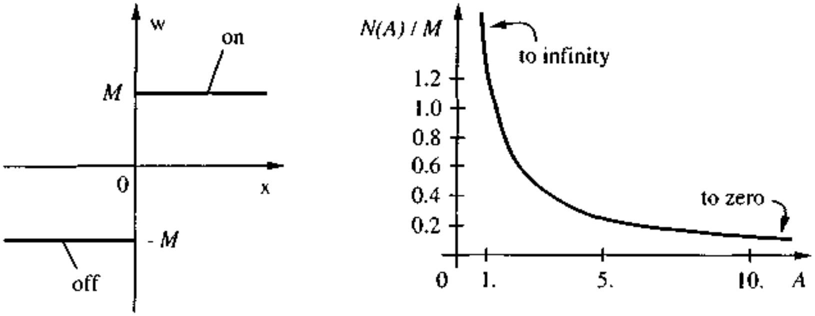

Figure 5.13 : Relay nonlinearity and its describing function

DEAD-ZONE

Consider the dead-zone characteristics shown in Figure 5.9, with the dead-zone width being 2δ and its slope k. The response corresponding to a sinusoidal input $x(t) = A\sin(\omega t)$ into a dead-zone of width 2δ and slope k, with $A \geq \delta$, is plotted in Figure 5.14. Since the characteristics is an odd function, $a_1 = 0$. The response is also seen to be symmetric over the four quarters of a period. In one quarter of a period, *i.e.*, when $0 \leq \omega t \leq \pi/2$, one has

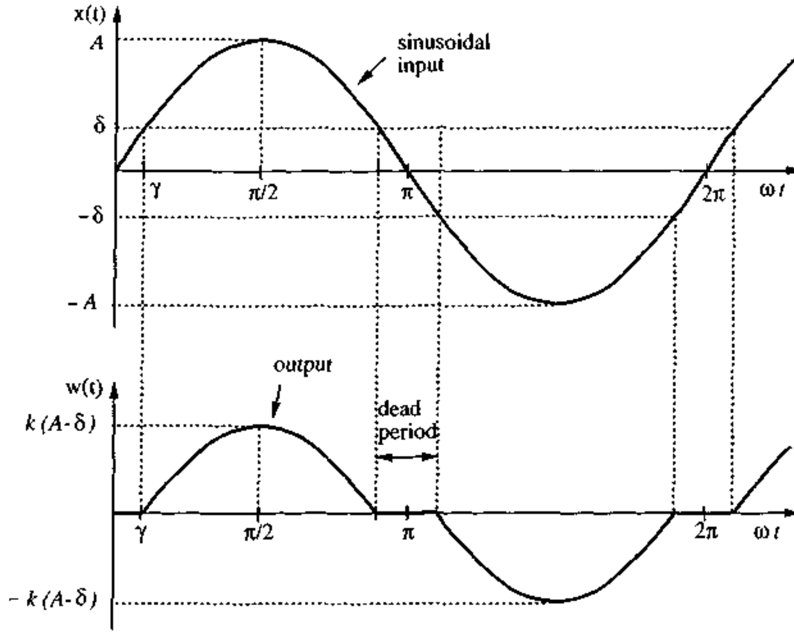

Figure 5.14 : Input and output functions for a dead-zone nonlinearity

$$w(t) = \begin{cases} 0 & 0 \leq \omega t \leq \gamma \\ k(A\sin(\omega t) - \delta) & \gamma \leq \omega t \leq \pi/2 \end{cases}$$

where $\gamma = \sin^{-1}(\delta/A)$. The coefficient b_1 can be computed as follows

$$b_1 = \frac{4}{\pi}\int_0^{\pi/2} w(t)\sin(\omega t)\, d(\omega t) \;=\; \frac{4}{\pi}\int_\gamma^{\pi/2} k(A\sin(\omega t) - \delta)\sin(\omega t)\, d(\omega t)$$

$$= \frac{2kA}{\pi}\left(\frac{\pi}{2} - \sin^{-1}\frac{\delta}{A} - \frac{\delta}{A}\sqrt{1 - \frac{\delta^2}{A^2}}\,\right) \tag{5.13}$$

This leads to

$$N(A) = \frac{2k}{\pi}\left(\frac{\pi}{2} - \sin^{-1}\frac{\delta}{A} - \frac{\delta}{A}\sqrt{1 - \frac{\delta^2}{A^2}}\,\right)$$

This describing function $N(A)$ is a *real* function and, therefore, there is no phase shift (reflecting the absence of time-delay). The normalized describing function is plotted in Figure 5.15. It is seen that $N(A)/k$ is zero when $A/\delta < 1$, and increases up to 1 with A/δ. This increase indicates that the effect of the dead-zone gradually diminishes as the amplitude of the input signal is increased, consistently with intuition.

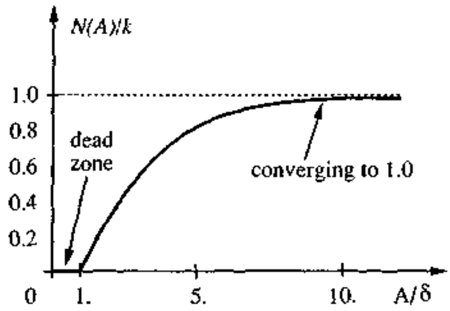

Figure 5.15 : Describing function of the dead-zone nonlinearity

BACKLASH

The evaluation of the describing functions for backlash nonlinearity is more tedious. Figure 5.16 shows a backlash nonlinearity, with slope k and width $2b$. If the input amplitude is smaller than b, there is no output. In the following, let us consider the input being $x(t) = A\sin(\omega t)$, $A \geq b$. The output $w(t)$ of the nonlinearity is as shown in the figure. In one cycle, the function $w(t)$ can be represented as

$$w(t) = (A - b)k \qquad \frac{\pi}{2} < \omega t \leq \pi - \gamma$$

$$w(t) = (A\sin(\omega t) + b)k \qquad \pi - \gamma < \omega t \leq \frac{3\pi}{2}$$

$$w(t) = -(A - b)k \qquad \frac{3\pi}{2} < \omega t \leq 2\pi - \gamma$$

$$w(t) = (A\sin(\omega t) - b)k \qquad 2\pi - \gamma < \omega t \leq \frac{5\pi}{2}$$

where $\gamma = \sin^{-1}(1 - 2b/A)$.

Unlike the previous nonlinearities, the function $w(t)$ here is neither odd nor even. Therefore, a_1 and b_1 are both nonzero. Using (5.7b) and (5.7c), we find through some tedious integrations that

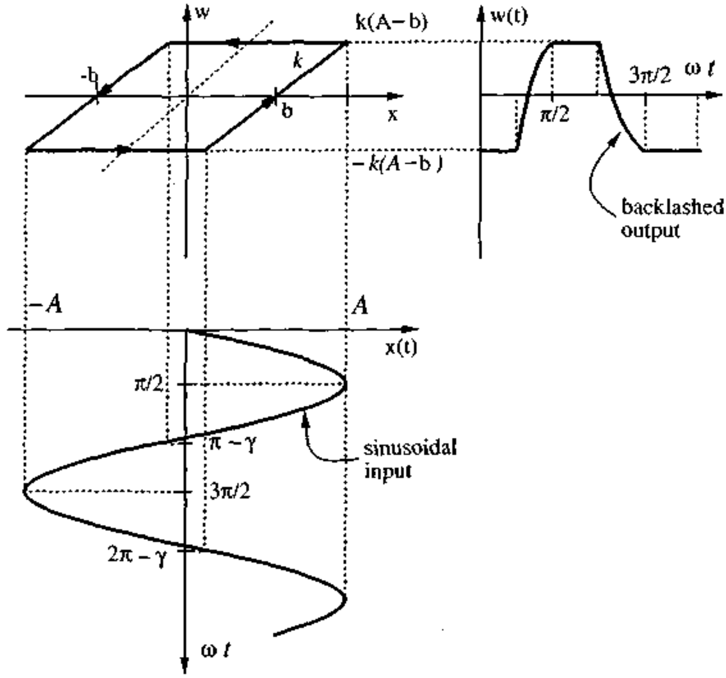

Figure 5.16 : Input and output functions for a backlash nonlinearity

$$a_1 = \frac{4kb}{\pi}\left(\frac{b}{A} - 1\right)$$

$$b_1 = \frac{Ak}{\pi}\left[\frac{\pi}{2} - \sin^{-1}\left(\frac{2b}{A} - 1\right) - \left(\frac{2b}{A} - 1\right)\sqrt{1 - \left(\frac{2b}{A} - 1\right)^2}\right]$$

Therefore, the describing function of the backlash is given by

$$|N(A)| = \frac{1}{A}\sqrt{{a_1}^2 + {b_1}^2} \tag{5.14a}$$

$$\underline{/N(A)} = \tan^{-1}(a_1/b_1) \tag{5.14b}$$

The amplitude of the describing function for backlash is plotted in Figure 5.17.

We note a few interesting points :

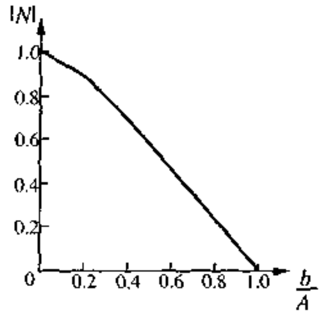

Figure 5.17 : Amplitude of describing function for backlash

1. $|N(A)| = 0 \quad if \; A = b$.
2. $|N(A)|$ increases, when b/A decreases.
3. $|N(A)| \to 1 \quad as \; b/A \to 0$.

The phase angle of the describing function is plotted in Figure 5.18. Note that a phase lag (up to 90^o) is introduced, unlike the previous nonlinearities. This phase lag is the reflection of the time delay of the backlash, which is due to the gap b. Of course, a larger b leads to a larger phase lag, which may create stability problems in feedback control systems.

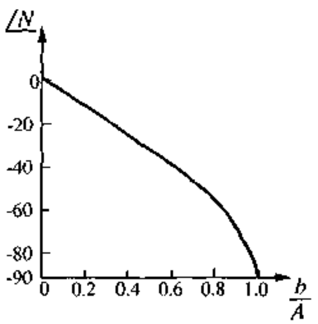

Figure 5.18 : Phase angle of describing function for backlash (degree)

5.4 Describing Function Analysis of Nonlinear Systems

For a nonlinear system containing a nonlinear element, we now know how to obtain a decribing function for the nonlinear element. The next step is to formalize the procedure in subsection 5.1.1 for the prediction of limit cycles, based on the describing function representation of the nonlinearity. The basic approach to achieve this is to apply an extended version of the famous Nyquist criterion in linear control to the equivalent system. Let us begin with a short review of the Nyquist criterion and its extension.

5.4.1 The Nyquist Criterion and Its Extension

Consider the linear system of Figure 5.19. The characteristic equation of this system is

$$\delta(p) = 1 + G(p)\,H(p) = 0$$

Note that $\delta(p)$, often called the *loop transfer function*, is a rational function of p, with its zeros being the poles of the closed-loop system, and its poles being the poles of the open-loop transfer function $G(p)\,H(p)$. Let us rewrite the characteristic equation as

$$G(p)\,H(p) = -1$$

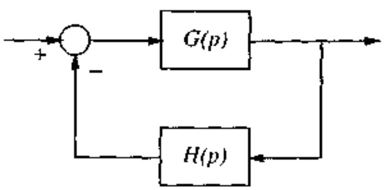

Figure 5.19 : Closed-loop linear system

Based on this equation, the famous Nyquist criterion can be derived straightforwardly from the Cauchy theorem in complex analysis. The criterion can be summarized (assuming that $G(p)\,H(p)$ has no poles or zeros on the $j\omega$ axis) in the following procedure (Figure 5.20):

1. draw, in the p plane, a so-called Nyquist path enclosing the right-half plane
2. map this path into another complex plane through $G(p)H(p)$
3. determine N, the number of clockwise encirclements of the plot of $G(p)H(p)$ around the point $(-1,0)$
4. compute Z, the number of zeros of the loop transfer function $\delta(p)$ in the right-half p plane, by

$$Z = N + P \quad , \text{ where } P \text{ is the number of unstable poles of } \delta(p)$$

Then the value of Z is the number of unstable poles of the closed-loop system.

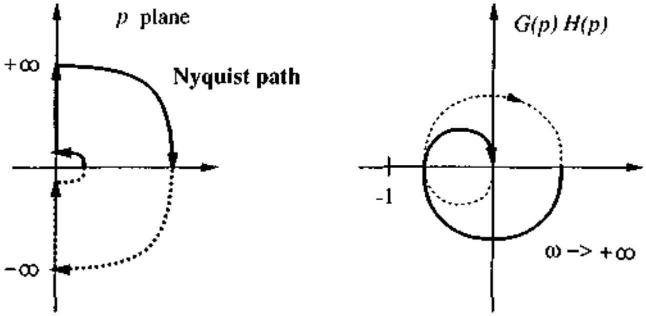

Figure 5.20 : The Nyquist criterion

A simple formal extension of the Nyquist criterion can be made to the case when a constant gain K (possibly a complex number) is included in the forward path in Figure 5.21. This modification will be useful in interpreting the stability analysis of limit cycles using the describing function method. The loop transfer function becomes

$$\delta(p) = 1 + K\, G(p) H(p)$$

with the corresponding characteristic equation

$$G(p)\, H(p) = -\, 1/K$$

The same arguments as used in the derivation of Nyquist criterion suggest the same procedure for determining unstable closed-loop poles, with the minor difference that now *Z represents the number of clockwise encirclements of the $G(p)\ H(p)$ plot around the point $-1/K$*. Figure 5.21 shows the corresponding extended Nyquist plot.

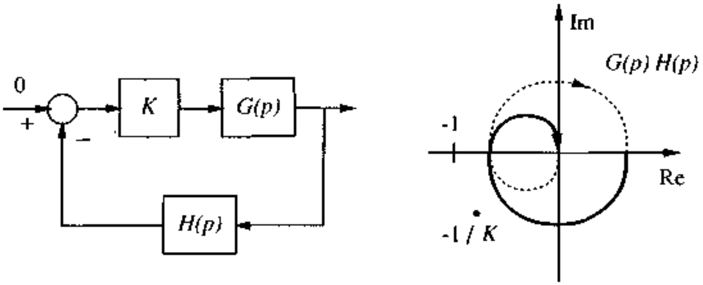

Figure 5.21 : Extension of the Nyquist criterion

5.4.2 Existence of Limit Cycles

Let us now assume that there exists a self-sustained oscillation of amplitude A and frequency ω in the system of Figure 5.22. Then the variables in the loop must satisfy the following relations:

$$x = -y$$

$$w = N(A,\omega)x$$

$$y = G(j\omega)\,w$$

Therefore, we have $y = G(j\omega)N(A,\omega)(-y)$. Because $y \neq 0$, this implies

$$G(j\omega)\,N(A,\omega) + 1 = 0 \tag{5.15}$$

which can be written as

$$G(j\omega) = -\frac{1}{N(A,\omega)} \tag{5.16}$$

Therefore, the amplitude A and frequency ω of the limit cycles in the system must satisfy (5.16). If the above equation has no solutions, then the nonlinear system has no limit cycles.

Expression (5.16) represents two nonlinear equations (the real part and imaginary part each give one equation) in the two variables A and ω. There are usually a finite number of solutions. It is generally very difficult to solve these equations by analytical methods, particularly for high-order systems, and therefore, a graphical approach is usually taken. The idea is to plot both sides of (5.16) in the complex plane and find the intersection points of the two curves.

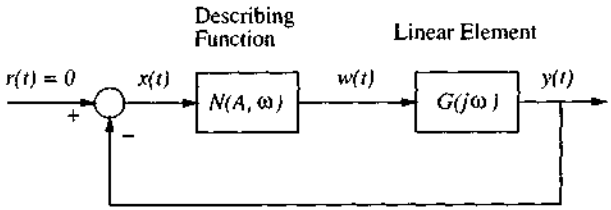

Figure 5.22 : A nonlinear system

FREQUENCY-INDEPENDENT DESCRIBING FUNCTION

First, we consider the simpler case when the describing function N being a function of the gain A only, *i.e.*, $N(A, \omega) = N(A)$. This includes all single-valued nonlinearities and some important double-valued nonlinearities such as backlash. The equality becomes

$$G(j\omega) = -\frac{1}{N(A)} \tag{5.17}$$

We can plot both the frequency response function $G(j\omega)$ (varying ω) and the negative inverse describing function $(-1/N(A))$ (varying A) in the complex plane, as in Figure 5.23. If the two curves intersect, then there exist limit cycles, and the values of A and ω corresponding to the intersection point are the solutions of Equation (5.17). If the curves intersect n times, then the system has n possible limit cycles. Which one is actually reached depends on the initial conditions. In Figure 5.23, the two curves intersect at one point K. This indicates that there is one limit cycle in the system. The amplitude of the limit cycle is A_k, the value of A corresponding to the point K on the $-1/N(A)$ curve. The frequency of the limit cycle is ω_k, the value of ω corresponding to the point K on the $G(j\omega)$ curve.

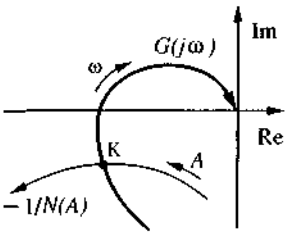

Figure 5.23 : Detection of limit cycles

Note that for single-valued nonlinearities, N is real and therefore the plot of $-1/N$ always lies on the real axis. It is also useful to point out that, as we shall discuss later, the above procedure only gives a *prediction* of the existence of limit cycles. The validity and accuracy of this prediction should be confirmed by computer simulations.

We already saw in section 5.1.1 an example of the prediction of limit cycles, for the Van der Pol equation.

FREQUENCY-DEPENDENT DESCRIBING FUNCTION

For the general case, where the describing function depends on both input amplitude and frequency ($N = N(A, \omega)$), the method can be applied, but with more complexity. Now the right-hand side of (5.15), $-1/N(A, \omega)$, corresponds to a family of curves on the complex plane with A as the running parameter and ω fixed for each curve, as shown in Figure 5.24. There are generally an *infinite* number of intersection points between the $G(j\omega)$ curve and the $-1/N(A, \omega)$ curves. Only the intersection points with matched ω indicate limit cycles.

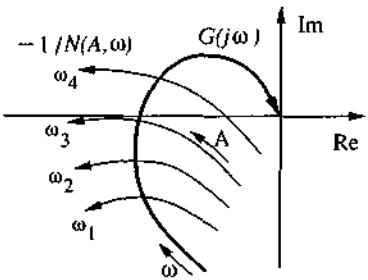

Figure 5.24 : Limit cycle detection for frequency-dependent describing functions

To avoid the complexity of matching frequencies at intersection points, it may be advantageous to consider the graphical solution of (5.16) directly, based on the plots of $G(j\omega)N(A, \omega)$. With A fixed and ω varying from 0 to ∞, we obtain a curve representing $G(j\omega)N(A,\omega)$. Different values of A correspond to a family of curves, as shown in Figure 5.25. A curve passing *through the point* $(-1,0)$ in the complex plane indicates the existence of a limit cycle, with the value of A for the curve being the amplitude of the limit cycle, and the value of ω at the point $(-1,0)$ being the frequency of the limit cycle. While this technique is much more straightforward than the previous one, it requires repetitive computation of the $G(j\omega)$ in generating the family of curves, which may be handled easily by computer.

5.4.3 Stability of Limit Cycles

As pointed out in chapter 2, limit cycles can be stable or unstable. In the above, we have discussed how to detect the existence of limit cycles. Let us now discuss how to determine the stability of a limit cycle, based on the extended Nyquist criterion in section 5.4.1.

Consider the plots of frequency response and inverse describing function in Figure 5.26. There are two intersection points in the figure, predicting that the system

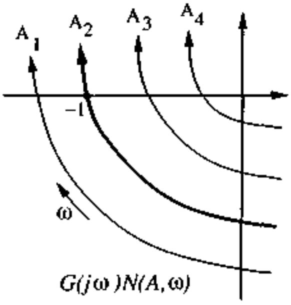

Figure 5.25 : Solving equation (5.15) graphically

has two limit cycles. Note that the value of A corresponding to point L_1 is smaller than the value of A corresponding to L_2. For simplicity of discussion, we assume that the linear transfer function $G(p)$ has no unstable poles.

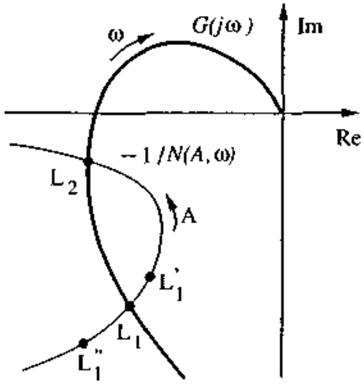

Figure 5.26 : Limit Cycle Stability

Let us first discuss the stability of the limit cycle at point L_1. Assume that the system initially operates at point L_1, with the limit cycle amplitude being A_1, and its frequency being ω_1. Due to a slight disturbance, the amplitude of the input to the nonlinear element is slightly increased, and the system operating point is moved from L_1 to L_1'. Since the new point L_1' is encircled by the curve of $G(j\omega)$, according to the extended Nyquist criterion mentioned in section 5.4.1, the system at this operating point is unstable, and the amplitudes of the system signals will increase. Therefore, the operating point will continue to move along the curve $-1/N(A)$ toward the other limit cycle point L_2. On the other hand, if the system is disturbed so that the amplitude A is decreased, with the operating point moved to the point L_1'', then A will continue to decrease and the operating point moving away from L_1 in the other direction. This is because L_1'' is not encircled by the curve $G(j\omega)$ and thus the extended Nyquist plot asserts the stability of the system. The above discussion indicates that a slight disturbance can destroy the oscillation at point L_1 and, therefore, that this limit cycle is unstable. A similar analysis for the limit cycle at point L_2 indicates that that limit cycle is stable.

Summarizing the above discussion and the result in the previous subsection, we obtain a criterion for existence and stability of limit cycles:

Limit Cycle Criterion: *Each intersection point of the curve $G(j\omega)$ and the curve $-1/N(A)$ corresponds to a limit cycle. If points near the intersection and along the increasing-A side of the curve $-1/N(A)$ are not encircled by the curve $G(j\omega)$, then the corresponding limit cycle is stable. Otherwise, the limit cycle is unstable.*

5.4.4 Reliability of Describing Function Analysis

Empirical evidence over the last three decades, and later theoretical justification, indicate that the describing function method can effectively solve a large number of practical control problems involving limit cycles. However, due to the approximate nature of the technique, it is not surprising that the analysis results are sometimes not very accurate. Three kinds of inaccuracies are possible:

1. The amplitude and frequency of the predicted limit cycle are not accurate
2. A predicted limit cycle does not actually exist
3. An existing limit cycle is not predicted

The first kind of inaccuracy is quite common. Generally, the predicted amplitude and frequency of a limit cycle always deviate somewhat from the true values. How much the predicted values differ from the true values depends on how well the nonlinear system satisfies the assumptions of the describing function method. In order to obtain accurate values of the predicted limit cycles, simulation of the nonlinear system is necessary.

The occurrence of the other two kinds of inaccuracy is less frequent but has more serious consequences. Usually, their occurrence can be detected by examining the linear element frequency response and the relative positions of the G plot and $-1/N$ plot.

Violation of filtering hypothesis: The validity of the describing function method relies on the filtering hypothesis defined by (5.5). For some linear elements, this hypothesis is not satisfied and errors may result in the describing function analysis. Indeed, a number of failed cases of describing function analysis occur in systems whose linear element has resonant peaks in its frequency response $G(j\omega)$.

Graphical Conditions: If the $G(j\omega)$ locus is tangent or almost tangent to the $-1/N$ locus, then the conclusions from a describing function analysis might be erroneous. Such an example is shown in Figure 5.27(a). This is because the effects of neglected higher harmonics or system model uncertainty may cause the change of the intersection situations, particularly when filtering in the linear element is weak. As a result, the second and third types of errors listed above may occur. A classic case of this problem involves a second-order servo with backlash studied by Nychols. While describing function analysis predicts two limit cycles (a stable one at high frequency and an unstable one at low frequency), it can be shown that the low-frequency unstable limit cycle does not exist.

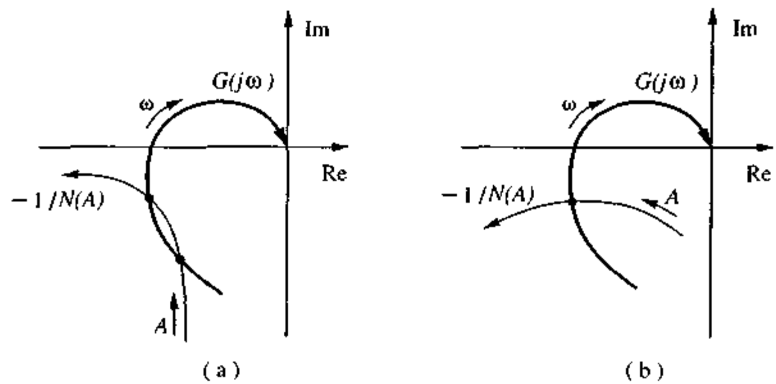

Figure 5.27 : Reliability of limit cycle prediction

Conversely, if the $-1/N$ locus intersects the G locus almost perpendicularly, then the results of the describing function are usually good. An example of this situation is shown in Figure 5.27(b).

5.5 Summary

The describing function method is an extension of the frequency response method of linear control. It can be used to *approximately* analyze and predict the behavior of important classes of nonlinear systems, including systems with hard nonlinearities. The desirable properties it inherits from the frequency response method, such as its graphical nature and the physically intuitive insights it can provide, make it an important tool for practicing engineers. Applications of the describing function method to the prediction of limit cycles were detailed. Other applications, such as predicting subharmonics, jump phenomena, and responses to external sinusoidal inputs, can be found in the literature.

5.6 Notes and References

An extensive and clear presentation of the describing function method can be found in [Gelb and VanderVelde, 1968]. A more recent treatment is contained in [Hedrick, *et al.*, 1982], which also discusses specific applications to nonlinear physical systems. The describing function method was developed and successfully used well before its mathematical justification was completely formalized [Bergen and Franks, 1971]. Figures 5.14 and 5.16 are adapted from [Shinners, 1978]. The Van der Pol oscillator example is adapted from [Hsu and Meyer, 1968].

5.7 Exercises

5.1 Determine whether the system in Figure 5.28 exhibits a self-sustained oscillation (a limit cycle). If so, determine the stability, frequency, and amplitude of the oscillation.

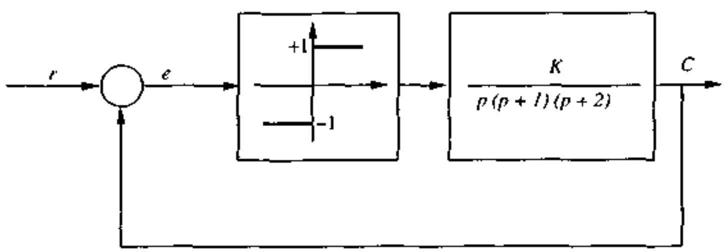

Figure 5.28 : A nonlinear system containing a relay

5.2 Determine whether the system in Figure 5.29 exhibits a self-sustained oscillation. If so, determine the stability, frequency, and amplitude of the oscillation.

5.3 Consider the nonlinear system of Figure 5.30. Determine the largest K which preserves the stability of the system. If $K = 2K_{max}$, find the amplitude and frequency of the self-sustained oscillation.

5.4 Consider the system of Figure 5.31, which is composed of a high-pass filter, a saturation function, and the inverse low-pass filter. Show that the system can be viewed as a nonlinear low-

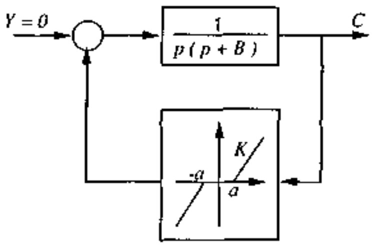

Figure 5.29 : A nonlinear system containing a dead-zone

pass filter, which attenuates high-frequency inputs *without introducing a phase lag.*

5.5 This exercise is based on a result of [Tsypkin, 1956].

Consider a nonlinear system whose output $w(t)$ is related to the input $u(t)$ by an odd function, of the form

$$w(t) = F(u(t)) = -F(-u(t)) \tag{5.18}$$

Derive the following very simple approximate formula for the describing function $N(A)$

$$N(A) \approx \frac{2}{3A}\,[\,F(A) + F(A/2)\,]$$

To this effect, you may want to use the fact that

$$\frac{1}{\pi}\int_{-1}^{1}\frac{f(x)}{\sqrt{1-x^2}}\,dx = \frac{1}{6}\,[\,f(1) + f(-1) + 2f(1/2) + 2f(-1/2)\,] + R$$

where the remainder R verifies $R = f^6(\xi)/(2^5\,6!)$ for some $\xi \in (-1, 1)$. Show that approximation (5.18) is quite precise (how precise?).

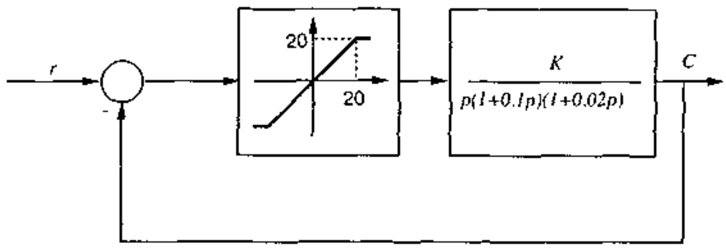

Figure 5.30 : A nonlinear system containing a saturation

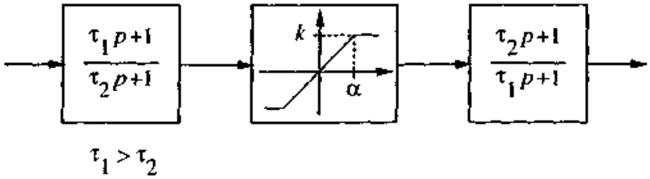

Figure 5.31 : A nonlinear low-pass filter

Invert (5.18) so as to obtain for the input-output relation a solution of the form

$$F(A) \approx \sum_{k=0}^{\infty} (-1)^k \, \frac{3A}{2^{k+1}} \, N\left(\frac{A}{2^k}\right)$$

5.6 In this exercise, adapted form [Phillips and Harbor, 1988], let us consider the system of Figure 5.32, which is typical of the dynamics of electronic oscillators used in laboratories, with

$$G(p) = \frac{-5p}{p^2 + p + 25}$$

Use describing function analysis to predict whether the system exhibits a limit cycle, depending on the value of the saturation level k. In such cases, determine the limit cycle's frequency and amplitude.

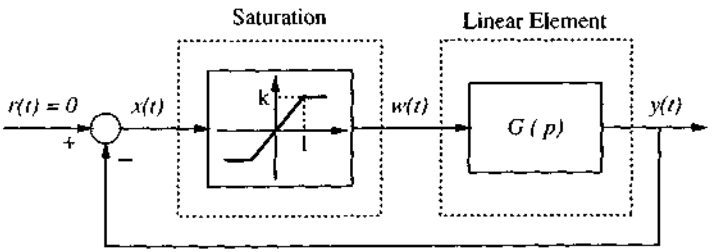

Figure 5.32 : Dynamics of an electronic oscillator

Interpret intuitively, by assuming that the system is started at some small initial state, and noticing that $y(t)$ can stay neither at small values (because of instability) nor at saturation values (by applying the final value theorem of linear control).