NONLINEAR CONTROL SYSTEMS, ANALYTICAL METHODS

A large number of methods exist for the analysis and design of linear control systems. Unlike a linear system, a nonlinear system does not satisfy the superposition property, which means not only that it may be difficult to deduce how the system will respond to a specific input if its response is known to a different input, but also nonlinear systems exhibit unique behavior due to the effects of the nonlinearity. In many systems it is possible to separate the static nonlinear effects from the dynamic so that a nonlinear system can often be accurately modeled as a combination of static nonlinear elements and linear dynamic elements. Here the concentration is primarily on analytical methods for nonlinear systems which are associated with those aspects of linear theory usually referred to as classical control. This means basically that for systems other than second order, frequency domain, rather than state space, models and formulations are used. An exception to this is the material on variable structure systems. The state space theme, and some design methods presented within that framework are given in the following article.

A block diagram of a simple nonlinear feedback system which will receive significant attention in this article is shown in Fig. 1. Although it only contains one nonlinear element, its presence can change the whole behavioral possibilities of the feedback loop compared with the linear situation, and its form is adequate for discussing many of the analysis and design techniques presented in this article. Since all practical systems contain some form of nonlinearity, it is important that basic concepts relating to the effects of nonlinearity are well understood. When this is the case it will allow the designer to assess qualitatively, if not quantitatively, the possible effects of nonlinear operation at various points within a feedback system and to take them into account in the design. This may allow the analysis and design to be done using one or more linearized models. A full nonlinear simulation may then be used to check that the design works satisfactorily when the nonlinear effects are included. This approach works satisfactorily in many instances, particularly if gain scheduling is used to counteract the effects of changes produced by any nonlinearity; however, this approach cannot be used for all situations.

Many nonlinear effects which take place in control systems may be modeled approximately using static nonlinearities. These include saturation in amplifiers, dead zones in valves, friction, and backlash in gears. Depending on the approach to be used in the analysis or design when these nonlinearities exist, it may be necessary to approximate their characteristics by either simple continuous mathematical functions, such as polynomials, or linear segmented approximations. To apply some methods which we will discuss, it may even be necessary to use coarser approximations to a nonlinearity simply to say that it is confined within a sector. These types of nonlinearities are often referred to as inherent nonlinearities, since for a satisfactory design they will exist due to the devices used, although for analysis we may wish to neglect them. It is also true that good designs will always be nonlinear, since trying to ensure linear operation of a system will involve the selection of oversized components such as pumps, motors, and heaters. Nonlinearity may be introduced intentionally into control systems to compensate for existing nonlinear effects, or to implement a design strategy which is either preferable technically or more economical.

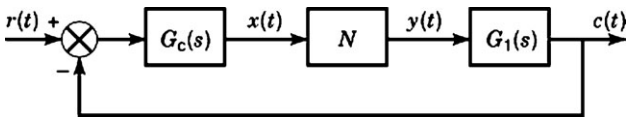

Figure 1. A simple nonlinear feedback system.

The simple feedback system of Fig. 1, provided that the nonlinearity or transfer functions are suitably chosen, may exhibit a variety of behaviors which are unique to nonlinear systems. First, the performance of the system, even for a specific type of input, will depend upon the amplitude of the input. The response, for example, to a small step input may be quite different from that of a large step input. If the autonomous system—that is, the system with no input—is released from several initial states, then the resulting behavior may be appreciably different for each state. For example, instead of reaching a stationary equilibrium, the system may move from some initial conditions into a limit cycle, a continuous oscillation which can be reached from a subset of initial conditions. This behavior is distinct from an oscillation in an idealized linear system since the magnitude of this latter oscillation is dependent upon the initial energy input to the system. A limit cycle is a periodic motion, but its waveform may be significantly different from the sinusoid of an oscillation. The autonomous nonlinear system may also have a chaotic motion, a motion which is repeatable from given initial conditions but which exhibits no easily describable mathematical form, is not periodic, and exhibits a spectrum of frequency components.

If a sinusoidal input is applied to the system, then the output may be of the same frequency but will also contain harmonics or other components related to the input frequency. This output too, for certain frequencies and amplitudes of the input, may not be unique but has an amplitude dependent upon the past history of the input or the initial conditions of the system. The sinusoidal input may also cause the system to oscillate at a related frequency so that the largest frequency component in the output is not the same as that of the input. Also if, for example, the autonomous system has a limit cycle, then the addition of a sinusoidal input will cause the limit cycle frequency to change and possibly cause synchronization of the limit cycle frequency with the input frequency or one of its harmonics. In many instances the phenomena just mentioned are undesirable in a control system, so that one needs techniques to ensure that they do not occur. Control systems must be designed to meet specific performance objectives, and to do this one is required to design a control law which is implemented based on measurements or estimation of the system states or, by simple functions of the system variables, typically the error signal. Many systems can be

made to operate satisfactorily with the addition of a simple controller in the error channel, which is shown by the transfer function $G_c(s)$ in Fig. 1. Typical performance criteria, which the system may be required to meet, are that it is stable, has zero steady-state error and a good response to a step input, suitably rejects disturbances, and is robust to parameter variations. Although one reason for using feedback control is to reduce sensitivity to parameter changes, specific design techniques can be used to ensure that the system is more robust to any parameter changes. If the process to be controlled is strongly nonlinear, then a nonlinear controller will have to be used if it is required to have essentially the same step response performance for different input step amplitudes. Some control systems—for example, simple temperature control systems—may work in a limit cycle mode, so that in these instances the designer is required to ensure that the frequency and amplitude variations of the controlled temperature are within the required specifications.

In the next section, we examine in some detail various approaches which can be used for investigating the analysis and design of nonlinear systems. The first topic discussed is the phase plane method, which can normally only be used to investigate second-order systems. It is a useful technique, since it can be used when more than one nonlinearity exists in the system and since many control problems, such as position control systems, can be modeled approximately by second-order dynamics. As mentioned previously, one specification for the design may be that the system must be stable. For linear systems, assessment of stability is a simple problem, but this is not the case for a nonlinear system, even when it is as simple as that shown in Fig. 1. Several absolute stability criteria exist for checking whether the system of Fig. 1 will be stable, and these are discussed in more detail later. The criteria presented are easy to use; and the circle criterion in particular, being basically an extension of the Nyquist criterion, is easy to implement and follow. A disadvantage, however, is that all these criteria only produce sufficient conditions so that if the condition is violated the system may still be stable.

To try to obtain an estimate of the possibility of this being the situation, the describing function method has been used by engineers for many years. The difficulty with the describing function approach, which approximates the nonlinearity by its gain to a sinusoidal input, is that the results are not exact. It does, however, enable the designer to obtain more insight into the situation, and, of course, the ideas can often be further checked by simulation. The describing function approach can also be helpful for system design in terms of shaping the frequency response of the system to produce a more stable situation or for indicating possible nonlinear effects which can be added in the controller to counteract the nonlinear effects in the plant. Describing functions for other than a single sinusoid can be obtained, and these allow some of the more complex aspects of the behavior of nonlinear systems to be investigated. These include, for example, synchronization and subharmonic generation as well as estimating more accurately the frequency content of any limit cycle. Relay-type characteristics are often introduced in control system to provide economic designs or to produce variable structure systems. First, a method for the determination of limit cycles in relay systems is presented. This is an interesting approach, since it allows the exact evaluation of a limit cycle and also an exact determination of whether it is stable or not. The method in this sense is unique, since exact limit cycle data for systems with any order dynamics containing a relay can be obtained.

As with the design of linear control systems, the issue of robustness to unmodeled dynamics and parameter uncertainty is also pertinent in the nonlinear control area. One such robust technique is the so-called variable structure or sliding mode approach. Variable structure control systems (*VSCS*) are characterized by a set of feedback control laws and an associated decision rule or switching function. This decision rule has as its input some measure of the current system behavior and produces as an output the particular feedback control law which should be used at that instant in time. The resulting variable structure system (*VSS*) may be regarded as a combination of subsystems, where each subsystem has a fixed control law which is valid for specified regions of system behavior. One of the advantages of introducing this additional complexity into the system is the ability to combine useful properties of each of the subsystems into a very flexible closed-loop control strategy. Indeed, it will be seen that a VSS may be designed to possess new properties which are not present in any of the composite structures. Utilization of these natural ideas began in the late 1950s in the Soviet Union and formed the foundations for significant contributions to the area of robust nonlinear control.

Of particular interest in the area of VSS is the so-called sliding mode behavior, where the control is designed to drive and then constrain the system state and lie within a neighborhood of the switching function. There are two significant advantages with this approach to controller design. First, the dynamic behavior of the system may be tailored by the particular choice of switching function. Second, the closed-loop response becomes totally insensitive to changes in certain plant parameters and will completely reject a particular class of external disturbances. This invariance property clearly renders sliding mode control a strong candidate for robust control. In addition, the ability to specify performance directly makes sliding mode control attractive from the design perspective. This is seen from the wide exposure of sliding mode control to many applications areas including robotics, aerospace, and automotive industries.

The sliding mode design approach involves two stages. The first consists of the design of an appropriate switching function to ensure that the system behavior during sliding motion satisfies the required design specifications. This is termed the *existence problem.* In the simplest case, this will be seen to amount to the design of a linear full-state feedback controller for a particular subsystem. The second design stage is concerned with the selection of a control law which will make the switching function attractive to the system state. This is termed the *reachability problem.* It is important to note that this control law is not necessarily discontinuous in nature.

THE PHASE PLANE METHOD

The phase plane method was the first approach used by control engineers for studying the effects of nonlinearity in feedback control systems. The technique can generally only be used for systems represented by second-order differential equations. It had previously been used in nonlinear mechanics and for the study of nonlinear oscillations. Smooth mathematical functions were assumed for the nonlinearities so that the second-order equation could be represented by two nonlinear first-order equations of the form

$$\begin{aligned} \dot{x}_1 &= P(x_1, x_2) \\ \dot{x}_2 &= Q(x_1, x_2) \end{aligned} \tag{1}$$

Equilibrium, or singular points, occur when

$$\dot{x}_1 = \dot{x}_2 = 0$$

and the slope of any solution curve, or trajectory, in the $x_1 - x_2$ state plane is

$$\frac{dx_2}{dx_1} = \frac{\dot{x}_2}{\dot{x}_1} = \frac{Q(x_1, x_2)}{P(x_1, x_2)} \tag{2}$$

A second-order nonlinear differential equation representing a control system with smooth nonlinearity can typically be written as

$$\ddot{x} + f(x, \dot{x}) = 0$$

and if this is rearranged as two first-order equations, choosing the phase variables as the state variables—that is, $x_1 = x$, $x_2 = x^2$—it can be written as

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -f(x_1, x_2) \end{aligned} \tag{3}$$

which is a special case of Eq. (3). A variety of procedures have been proposed for sketching state (phase) plane trajectories for Eqs. 3 and 5. A complete plot showing trajectory motions throughout the entire state (phase) plane from different initial conditions is known as a state (phase) portrait. Knowledge of these original methods, despite the immense improvements in computation since they were first proposed, can be particularly helpful for obtaining an appreciation of the system behavior. When simulation studies are undertaken, phase plane graphs are easily obtained and they are often more helpful for understanding the system behavior than displays of the variables x_1 and x_2 against time.

Many investigations using the phase plane technique were concerned with the possibility of limit cycles in the nonlinear differential equations. When a limit cycle exists, this results in a closed trajectory in the phase plane; typical of such investigations was the work of Van der Pol. He considered the equation

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

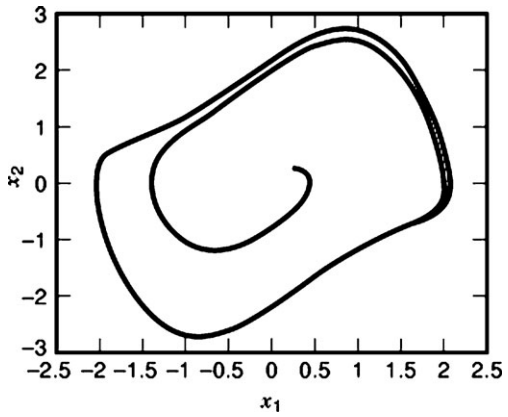

Figure 2. Limit cycle solution of Van der Pol equation with $\mu = 1.0$.

where μ is a positive constant. The phase plane form of this equation can be written as

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -f(x_1, x_2) = \mu(1 - x_1^2)x_2 - x_1 \end{aligned}$$

The slope of a trajectory in the phase plane is

$$\frac{dx_2}{dx_1} = \frac{\dot{x}_2}{\dot{x}_1} = \frac{\mu(1 - x_1^2)x_2 - x_1}{x_2} \tag{4}$$

which is only singular (that is, at an equilibrium point), when the right-hand side of Eq. (4) is 0/0, that is, $x_1 = x_2 = 0$.

The form of the singular point, which is obtained from linearization of the equation at the origin, depends upon μ, being an unstable focus for $\mu < 2$ and an unstable node for $\mu > 2$. All phase plane trajectories have a slope of r when they intersect the curve

$$rx_2 = \mu(1 - x_1^2)x_2 - x_1 \tag{5}$$

One way of sketching phase plane behavior is to draw a set of curves for various selected values of r in Eq. (5) and marking the trajectory slope r on the curves, a procedure known as the method of isoclines. Figure 2 shows a simulation result from a small initial condition leading to the stable limit cycle solution for $\mu = 1.0$.

Many nonlinear effects in control systems, such as saturation and friction, are best approximated by linear segmented characteristics rather than continuous mathematical functions. This is an advantage for study using the phase plane approach, since it results in a phase plane divided up into different regions but with a linear differential equation describing the motion in each region.

To illustrate the approach, consider a basic relay position-control system with nonlinear velocity feedback having the block diagram shown in Fig. 3. First, let us assume that the hysteresis in the relay is negligible (i.e., $\Delta = 0$) and that h is large so that the velocity-feedback signal will not saturate. Denoting the system position output by x_1 and its derivative x^2_1 by x_2, we note that the relay output of ± 1 or 0 is equal to $\overset{22}{x}_{1/K}$ and that the relay input is equal to $-x_1 - \lambda x_2 = -1$. Taking the dead zone of the relay

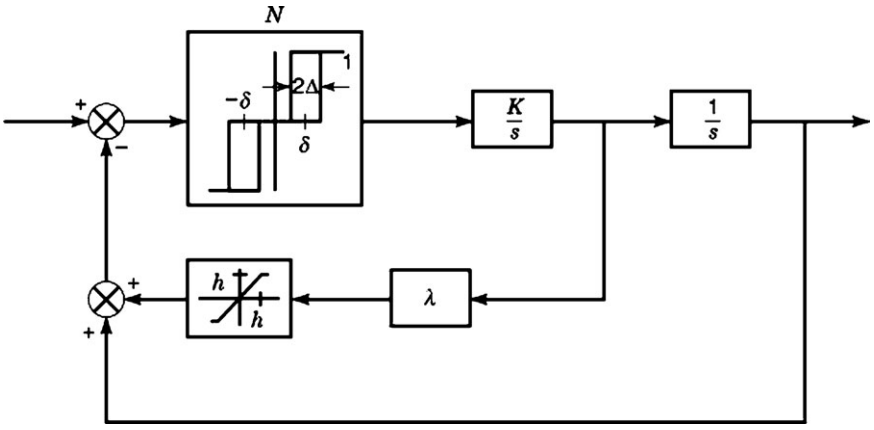

Figure 3. Block diagram of relay position-control system.

$\pm\delta$ to be equal to ± 1, the motion of the system is described by

$$\ddot{x} = \begin{cases} K & \text{if } -x_1 - \lambda x_2 > 1 \\ 0 & \text{if } |-x - \lambda x_2| < 1 \\ -K & \text{if } -x_1 - \lambda x_2 = -1 \end{cases}$$

Thus in the phase plane, which has x_1 as abscissa and x_2 as ordinate, the dashed lines $x_1 + \lambda x_2 = \pm 1$ in Fig. 4 divide the plane into the three regions, for each of which the motion is described by one of the above three simple linear second-order differential equations. The solution of

$$\ddot{x}_1 = K$$

in terms of the phase-plane coordinates x_1 and x_2 is

$$x_2^2 - x_{20}^2 = 2K(x_1 - x_{10}) \tag{6}$$

where x_{10} and x_{20} are the initial values of x_1 and x_2. Since Eq. (6) describes a parabola, which for the special case of $K = 0$ has the solution $x_2 = x_{20}$, it is easy to calculate the system's response from any initial condition (x_{10}, x_{20}) in the phase plane. Figure 4 shows the response from $(-4.6, 0)$ with $\lambda = 1$ and $K = 1.25$. The initial parabola meets the first switching boundary at A; the ensuing motion is horizontal—that is, at constant velocity—until the second switching boundary is reached at B. The resulting parabola meets the same switching boundary again at C, at which point motion from either side of the switching line through C will be directed toward C, so that the resulting motion is a sliding motion. Responses from any other initial conditions are obviously easy to find, but, from the one response shown, several aspects of the system's behavior are readily apparent. In particular, the system is seen to be stable since all responses will move inward, possibly with several overshoots and undershoots, and will finally slide down a switching boundary to ± 1. Thus a steady-state error of unit magnitude will result from any motion.

When the velocity-feedback signal saturates—that is, when $|\lambda x_2| > h$—the input signal to the relay is $-x_1 \pm h$. The switching boundaries change to those shown in Fig. 5, but the equations describing the motion between the boundaries remain unaltered. Therefore for a large step input the response will become more oscillatory when the velocity saturates. When the hysteresis is finite then the switching lines for positive (negative) x_2 move to the right(left) at their intersection with the x_1 axis. If h is large it is then easily shown that a limit cycle, as shown in Fig. 6 for $\delta = 1$ and $\Delta = 0.5$, will occur. Trajectories both inside and outside the limit cycle have their motion directed toward it. Similarly, it is straightforward to draw phase-plane trajectories for a finite hysteresis Δ and smaller values of h.

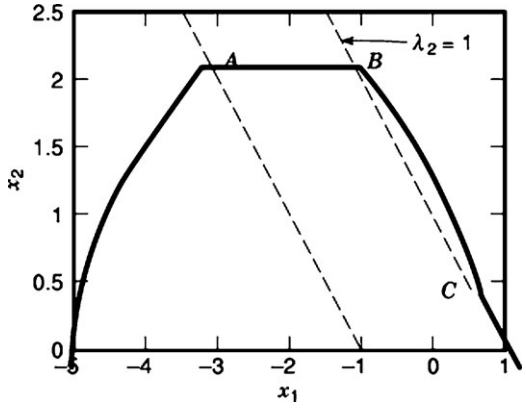

Figure 4. Initial condition response.

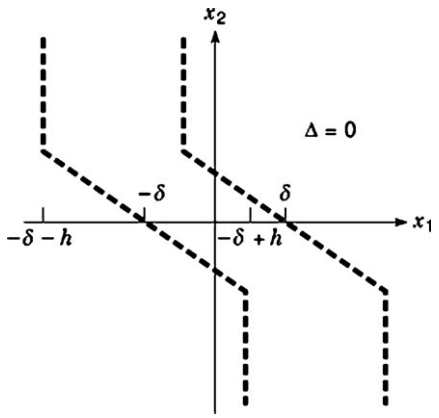

Figure 5. Changed switching boundaries due to saturation.

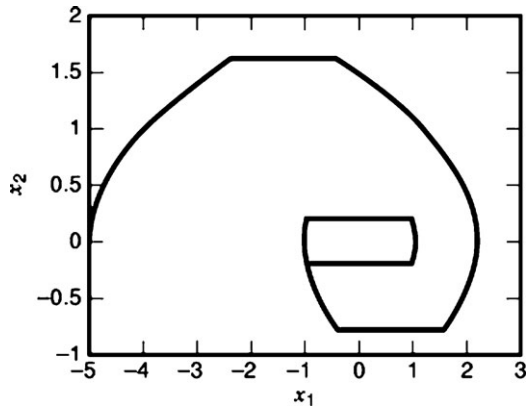

Figure 6. Response terminating in a limit cycle for $\delta = 1$ and $\Delta = 0.5$.

ABSOLUTE STABILITY CRITERIA

A very important question in control is to be able to ascertain the stability of a feedback system. The problem for linear systems was examined over a century ago in Cambridge, England, by Routh, who published his famous work on the stability of motion in 1877. As a result of this work and further contributions, most notably by Nyquist, several approaches are now available for determining the stability of a feedback loop such as Fig. 1 when the nonlinearity $n(x)$ is replaced by a linear gain K. The methods provide necessary and sufficient conditions for stability. The first work on the stability of nonlinear systems by Lyapunov was published in 1892, and since that time there have been many attempts to determine necessary and sufficient conditions for the stability of the autonomous feedback system—that is, $r = 0$—of Fig. 1. Lyapunov formulated an approach for determining sufficient conditions, but the difficulty of his method is that it requires determination of a function of the system states which then must be shown to satisfy certain properties. There is no general approach for finding a suitable function; when one is found, it does not guarantee that a "better" function does not exist which will prove stability in a larger domain in the state space. The problem has therefore been researched by many people with the objective of obtaining conditions for stability which are relatively easy to use.

Several frequency-domain results (1) giving sufficient, but not necessary, conditions for stability have been determined which use limited information about the nonlinearity, $n(x)$, typically its sector bounds or the sector bounds of its slope. The nonlinearity $n(x)$ has sector bounds (k_1, k_2); that is, it is confined between the straight lines k_1x and k_2x if $k_1x2 < xn(x) < k_2x^2$ for all x. Similarly, it has slope bounds (k'_1, k'_2) if $k'_1x^2 < xn'(x) < k'_2x^2$, where $n'(x) = dn(x)/dx$. The Popov criterion (2) states that a sufficient condition for the autonomous system of Fig. 1 to be stable if $G(s)$ is stable and $G(\infty) > -k{-}1$ is that a real number $q > 0$ can be found such that for all ω we obtain

$$\mathbf{Re}[(1 + j\omega q)G(j\omega)] + k^{-1} > 0$$

where the nonlinearity $n(x)$ lies in the sector $(0, k)$. The theorem has the simple graphical interpretation shown in

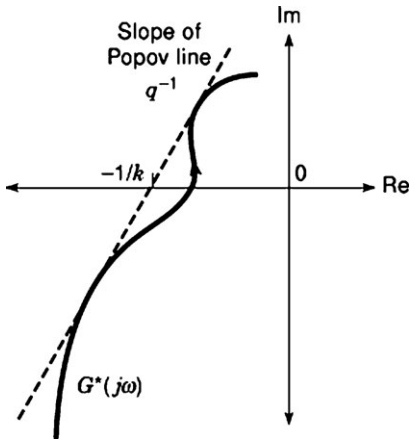

Figure 7. Graphical illustration of the Popov criterion.

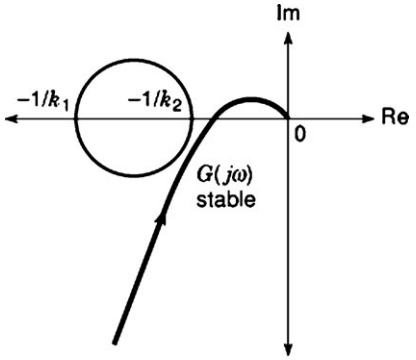

Figure 8. Illustration of the circle criterion.

Fig. 7, where for the system to be stable a line of slope q^{-1} can be drawn through the point $-k^{-1}$ so that the Popov locus $G^*(j\omega)$ lies to the right of the line. The Popov locus is given by

$$G^*(j\omega) = \mathbf{Re}[G(j\omega)] + j\omega\,\mathbf{Im}[G(j\omega)]$$

The circle criterion (3) is more conservative than the Popov criterion but can be used when both the nonlinearity is time varying and there is a bounded input, r, to the system of Fig. 1. Satisfaction of the circle criterion guarantees that the autonomous system is absolutely stable and the system with bounded input has a bounded output. The criterion uses the Nyquist locus, $G(j\omega)$, and for stability of the system of Fig. 1 with $n(x)$ in the sector (k_1, k_2) it is required that $G(j\omega)$ for all real ω has the following properties. If the circle C has its diameter from $-1/k_1$ to $-1/k_2$ on the negative real axis of the Nyquist diagram, then (1) if $k_1k_2 < 0$, $G(j\omega)$ should be entirely within C, (2) if $k_1k_2 > 0$, $G(j\omega)$ should lie entirely outside and not encircle C, and (3) if $k_1 = 0$ or $k_2 = 0$, $G(j\omega)$ lies entirely to the right of $-1/k_2$ or to the left of $-1/k_1$. The situation for stability in case (2) is shown in Fig. 8.

Two simple transformations are also useful for investigating the absolute stability of the autonomous system of Fig. 1. Feeding forward around the nonlinearity and backward around the dynamics $G(s)$, through a constant gain ρ, whose effects cancel out, changes the nonlinearity sector to $(k_1 - \rho, k_2 - \rho)$ and the linear transfer function

to $G(s)/[1+\rho G(s)]$. Alternatively feeding backward around $n(x)$ and forward around $G(s)$ changes the nonlinearity sector to $(k_1/(1-k_1\rho), (k_2/(1-k_2\rho))$ and changes the linear transfer function to $\rho + G(s)$. This is needed in order to apply the Popov criterion to the general finite sector—that is, $n(x)$ in the sector (k_1, k_2).

Prior to the derivation of these frequency-domain results, Aizermann had put forward a conjecture that the autonomous system of Fig. 1 would be stable for a nonlinearity sector bounded by (k_1, k_2) if for $k_1k_2 > 0$ the Nyquist locus $G(j\omega)$ of a stable transfer function did not touch or encircle the line between $-1/k_1$ and $-1/k_2$, which is of course the diameter of the circle of Fig. 8. Several counterexamples have been put forward to show that the conjecture is incorrect; however, it can be shown that if the conjecture is satisfied, the system may possess a limit cycle but its output cannot go unbounded (4). For a monotonic nonlinearity with slope bounds (k'_1, k'_2) and $k'_1k'_2 > 0$, an off-axis circle criterion exists (5). This states that the autonomous system of Fig. 1 with a nonlinearity satisfying the aforementioned conditions will be absolutely stable if the Nyquist locus of a stable transfer function does not encircle a circle centered off the real axis and which intercepts it at $(-1/k'_2, -1/k'_2)$.

DESCRIBING FUNCTION METHOD

The describing function (*DF*) method was developed simultaneously in several countries during the 1940s. Engineers found that control systems which were being used in many applications—for example, gun pointing and antenna control—could exhibit limit cycles under certain conditions rather than move to a static equilibrium. They realized that this instability was due to nonlinearities, such as backlash in the gears of the control system, and they wished to obtain a design method which could ensure that the resulting systems were free from limit cycle operation. They observed that when limit cycles occurred the observed waveforms at the system output were often approximately sinusoidal, and this indicated to them a possible analytical approach. Initial investigations therefore focused on the autonomous feedback system with a single nonlinear element shown in Fig. 1 containing a static nonlinearity $n(x)$ and linear dynamics given by the transfer function $G(s) = G_c(s)G_1(s)$. It was recognized that if a limit cycle existed in the autonomous system with the output $c(t)$ approximately sinusoidal, then the input $x(t)$ to the nonlinearity could be assumed sinusoidal, the corresponding fundamental output of the nonlinearity could be calculated, and conditions for this sinusoidal self-oscillation could be found, if the higher harmonics generated at the nonlinearity output were neglected. This is the concept of harmonic balance, in this case balancing the first harmonic only, which had previously been used by physicists to investigate such aspects as the generation of oscillations in electronic circuits. The DF of a nonlinearity was therefore defined as its gain to a sinusoid—that is, the ratio of the fundamental of the output to the amplitude of the sinusoidal input. Since describing functions can be used for other than a single sinusoidal input to a nonlinearity, as discussed in the latter part of this article; this DF is often, for clarity, called the sinusoidal DF (*SDF*).

The Sinusoidal Describing Function

We assume that if in Fig. 1 we have $x(t) = a\cos\theta$, where $\theta = \omega t$ and $n(x)$ is a symmetrical odd nonlinearity, then the output $y(t)$ will be given by the Fourier series.

$$y(\theta) = \sum_{n=0}^{\infty} a_n \cos n\theta + b_n \sin n\theta$$

where

$$a_0 = 0$$

$$a_1 = (1/\pi)\int_0^{2\pi} y(\theta)\cos\theta\, d\theta \tag{7}$$

and

$$b_1 = (1/\pi)\int_0^{2\pi} y(\theta)\sin\theta\, d\theta \tag{8}$$

The fundamental output from the nonlinearity is $a_1\cos\theta + b_1\sin\theta$, so that the DF is given by

$$N(a) = (a_1 - jb_1)/a$$

which may be written

$$N(a) = N_p(a) + jN_q(a)$$

where

$$N_p(a) = a_1/a \quad \text{and} \quad N_q(a) = -b_1/a$$

Alternatively, in polar coordinates,

$$N(a) = M(a)e^{j\Psi(a)}$$

where

$$M(a) = (a_1^2 + b_1^2)^{1/2}/a \quad \text{and} \quad \Psi(a) = -\tan^{-1}(b_1/a_1)$$

It is further easily shown that if $n(x)$ is single valued, then $b_1 = 0$. Although Eqs. 7 and 8 are an obvious approach to the evaluation of the fundamental output of a nonlinearity, they are somewhat indirect, in that one must first determine the output waveform $y(\theta)$ from the known nonlinear characteristic and sinusoidal input waveform. This is avoided if the substitution $\theta = \cos^{-1}(x/a)$ is made, in which case, after some simple manipulations, it can be shown that

$$a_1 = (4/a)\int_0^a x n_p(x) p(x)\, dx \tag{9}$$

$$b_1 = (4/a\pi)\int_0^a n_q(x)\, dx \tag{10}$$

The function $p(x)$ is the amplitude probability density function of the input sinusoidal signal and is given by

$$p(x) = (1/\pi)(a^2 - x^2)^{-1/2} \tag{11}$$

An additional advantage of using Eqs. 9 and q0 is that they easily yield proofs of some interesting properties of the DF for symmetrical odd nonlinearities. These include the following:

1. For a double-valued nonlinearity the quadrature component $N_q(a)$ is proportional to the area of the nonlinearity loop, that is, $N_q(a) = -(1/a^2\pi)$ (area of nonlinearity loop)
2. For two single-valued nonlinearities $n_\alpha(x)$ and $n_\beta(x)$, with $n_\alpha(x) < n_\beta(x)$ for all $0 < x < b$, we obtain $N_\alpha(a) < N_\beta(a)$ for input amplitudes less than b.
3. For the sector bounded single-valued nonlinearity—that is, $k_1 x < n(x) < k_2 x$ for all $0 < x < b$—we have $k_1 < N(a) < k_2$ for input amplitudes less than b. This is the sector property of the DF, and it also applies for a double-valued nonlinearity if $N(a)$ is replaced by $M(a)$.

When the nonlinearity is single-valued, it also follows directly from the properties of Fourier series that the DF, $N(a)$, may also be defined as follows:

1. The variable gain, K, having the same sinusoidal input as the nonlinearity, which minimizes the mean-squared value of the error between the output from the nonlinearity and that from the variable gain.
2. The covariance of the input sinusoid and the nonlinearity output divided by the variance of the input.

Evaluation of the Describing Function

Tables of DFs for a variety of nonlinear characteristics can be found in many books (6, 7). However, to illustrate the evaluation of the DF of a nonlinearity a few simple examples are considered below.

Cubic Nonlinearity. For this nonlinearity $n(x) = x^3$ and using Eq. (16), one has

$$\begin{aligned} a_1 &= (4/\pi)\int_0^{\pi/2} (a\cos\theta)^3 \cos\theta\, d\theta \\ &= (4/\pi)a^3 \int_0^{\pi/2} \cos^4\theta\, d\theta \\ &= (4/\pi)a^3 \int_0^{\pi/2} \left(\frac{3}{8} + \frac{\cos 2\theta}{2} + \frac{\cos 4\theta}{8}\right) d\theta = 3a^3/4 \end{aligned}$$

giving $N(a) = 3a^2/4$.

Alternatively from Eq. (23) we have

$$a_1 = (4/a)\int_0^a x^4 p(x)\, dx$$

The integral $\mu_n = \int_{-\infty}^{\infty} x^n p(x)\, dx$ is known as the nth moment of the probability density function; and for the sinusoidal distribution with $p(x) = (1/\pi)(a^2 - x^2)^{-1/2}$, μ_n has the value

$$\mu_n = \begin{cases} 0 & \text{for } n \text{ odd} \\ a^n \dfrac{(n-1)}{n}\dfrac{(n-3)}{(n-2)} \cdots \dfrac{1}{2} & \text{for } n \text{ even} \end{cases}$$

Therefore $a_1 = (4/a)\,½ \cdot ¾ \cdot ½\, a^4 = 3a^3/4$, as before.

Saturation Nonlinearity. The DF can also be found by taking the nonlinearity input as $a \sin\theta$, in which case for the ideal saturation characteristic shown in Fig. 9 the nonlinearity output waveform $y(\theta)$ is as shown in the same figure. Because of the symmetry of the nonlinearity the fundamental of the output can be evaluated from the integral over a quarter period so that

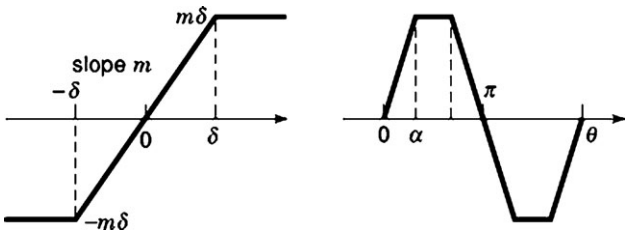

Figure 9. Saturation characteristic and output waveform.

$$N(a) = \frac{4}{a\pi}\int_0^{\pi/2} y(\theta)\sin\theta\, d\theta$$

which for $a > \delta$ gives

$$N(a) = \frac{4}{a\pi}\left[\int_0^{\alpha} ma\sin^2\theta\, d\theta + \int_{\alpha}^{\pi/2} m\delta\,\sin\theta\, d\theta\right]$$

with $\alpha = \sin^{-1}\delta/a$.

Evaluation of the integrals gives

$$N(a) = (4m/\pi)\left[\frac{\alpha}{2} - \frac{\sin 2\alpha}{4} + \delta\cos\alpha\right]$$

which on substituting for δ gives the result

$$N(a) = (m/\pi)(2\alpha + \sin 2\alpha)$$

Since for $a < \delta$ the characteristic is linear, giving $N(a) = m$, the DF for ideal saturation is $mN_S(\delta/a)$, where

$$N_S(\delta/a) = \begin{cases} 1 & \text{for } a < \delta \\ (1/\pi)[2\alpha + \sin 2\alpha] & \text{for } a > \delta \end{cases}$$

Alternatively one can integrate Eq. (9) by parts to give

$$a_1 = (4/a\pi)\int_0^a n'(x)(a^2 - x^2)^{1/2} dx \quad \text{if} \quad n(0) = 0$$

so that using the substitution $x = a\sin\theta$, this yields

$$N(a) = (4m/\pi)\int_0^{\alpha} \cos^2\theta\, d\theta = (m/\pi)(2\alpha + \sin 2\alpha)$$

as before.

Relay with Dead Zone and Hysteresis. The characteristic of a relay with dead zone and hysteresis is shown in Fig. 10 together with the corresponding input, assumed equal to $a\cos\theta$, and the corresponding output waveforms. Using

Eqs. 7 and 8 over the interval $-\pi/2$ to $\pi/2$ and assuming that the input amplitude a is greater than $\delta+\Delta$ gives

$$a_1 = (2/\pi)\int_{-\alpha}^{\beta} h\cos\theta\, d\theta = (2h/\pi)(\sin\beta+\sin\alpha)$$

where $\alpha=\cos^{-1}[(\delta-\Delta)/a]$ and $\beta=\cos^{-1}[(\delta+\Delta)/a,]$ and

$$b_1 = (2/\pi)\int_{-\alpha}^{\beta} h\sin\theta\, d\theta = (2h/\pi)\left(\frac{(\delta+\Delta)}{a}-\frac{(\delta-\Delta)}{a}\right) = 4h\Delta/a\pi$$

Thus

$$N(a)=\frac{2h}{a^2\pi}\{[a^2-(\delta+\Delta)^2]^{1/2}+[a^2-(\delta-\Delta)^2]^{1/2}\}-\frac{j4h\Delta}{a^2\pi} \quad (12)$$

For the alternative approach, one must first obtain the in-phase and quadrature nonlinearities which are shown in Fig. 11. Using Eqs. 23 and 24, one obtains

$$a_1 = (4/a)\int_{\delta-\Delta}^{\delta+\Delta} x(h/2)p(x)dx+\int_{\delta+\Delta}^{a} xh\,p(x)\,dx = \frac{2h}{a\pi}\{[a^2-(\delta+\Delta)^2]^{1/2}+[a^2-(\delta-\Delta)^2]^{1/2}\}$$

$$b_1 = (4/a\pi)\int_{d-\Delta}^{\delta+\Delta}(h/2)\,dx = 4h\Delta/a\pi = (\text{Area of nonlinearity loop})/a\pi$$

The DFs for other relay characteristics can easily be found from this result. For no hysteresis, $\Delta=0$; for no dead zone, $\delta=0$; and for an ideal relay, $\Delta=\delta=0$.

It is easily shown that the DF of two nonlinearities in parallel is equal to the sum of their individual DFs, a result which is very useful for determining DFs, particularly of linear segmented characteristics with multiple break points. Several procedures (6) are available for obtaining approximations for the DF of a given nonlinearity either by numerical integration or by evaluation of the DF of an approximating nonlinear characteristic defined, for example, by a quantized characteristic, linear segmented characteristic, or Fourier series.

Stability and Limit Cycles

To study the possibility of limit cycles in the autonomous closed loop system of Fig. 1, the nonlinearity $n(x)$ is replaced by its DF $N(a)$. Thus, the open-loop gain to a sinusoid is $N(a)G(j\omega)$ and a limit cycle will exist if

$$N(a)G(j\omega)=-1 \quad (13)$$

where $G(j\omega)=G_C(j\omega)G_1(j\omega)$. This condition means that the first harmonic is balanced around the closed loop. Since $G(j\omega)$ is a complex function of ω and $N(a)$ may be a complex function of a, a solution to Eq. (13) will yield both the frequency ω and amplitude a of a possible limit cycle.

Various approaches can be used to examine Eq. (13) with the choice affected to some extent by the problem—for example, whether the nonlinearity is single- or double-valued or whether $G(j\omega)$ is available from a transfer function $G(s)$ or as measured frequency response data. Typically the functions $G(j\omega)$ and $N(a)$ are plotted separately on Bode, Nyquist, or Nichols diagrams. Alternatively, stability criteria such as the Hurwitz–Routh or root locus plots may be used for the characteristic equation

$$1+N(a)G(s)=0$$

although here it should be remembered that the equation is appropriate only for $s\approx j\omega$.

Figure 12 illustrates the procedure on a Nyquist diagram, where the $G(j\omega)$ and $C(a)=-1/N(a)$ loci are plotted and shown intersecting for $a=a_0$ and $\omega=\omega_0$. The DF method therefore indicates that the system has a limit cycle with the input sinusoid to the nonlinearity, x, equal to $a_0\sin(\omega_0 t+\phi)$, where ϕ depends on the initial conditions. When the $G(j\omega)$ and $C(a)$ loci do not intersect, the DF method predicts that no limit cycle will exist if the Nyquist stability criterion is satisfied for $G(j\omega)$ with respect to any point on the $C(a)$ locus. Obviously, if the nonlinearity has unit gain for small inputs, the point $(-1,j0)$ will lie on $C(a)$ and it may then be used as the critical point, analogous to the situation for a linear system.

When the analysis indicates that the system is stable, its relative stability may be indicated by evaluating its gain and phase margin. These can be found for every amplitude a on the $C(a)$ locus, so it is usually appropriate to use the minimum values. In some cases a nonlinear block also includes dynamics so that its response is both amplitude and frequency dependent and its DF will be $N(a,\omega)$. A limit cycle will then exist if

$$G(j\omega)=-1/N(a,\omega)=C(a,\omega)$$

To check for possible solutions of this equation, a family of $C(a,\omega)$ loci, usually as functions of a for fixed values of ω, may be drawn on the Nyquist diagram.

A further point of interest when a solution to Eq. (40) exists is whether the predicted limit cycle is stable. This is obviously important if the control system is designed to have a limit cycle operation, as in the case of an on–off temperature control system, but it may also be important in other systems. If, for example, an unstable limit cycle condition is reached, the signal amplitudes will not become bounded but may continue to grow. The stability of a limit cycle, provided that only one solution is predicted, can be assessed by applying the Nyquist stability criterion to points on the $C(a)$ locus at both sides of the solution point. If the stability criterion indicates instability (stability) for a point on $C(a)$ with $a<a_0$ and indicates stability (instability) for a point on $C(a)$ with $a>a_0$, then the limit cycle is stable (unstable). The situation is more complicated when multiple limit cycle solutions exist and the above criterion is a necessary but not sufficient result for the stability of the limit cycle (8).

The stability of the limit cycle can then normally be ascertained by examining the roots of the characteristic equation

$$1+N_{i\gamma}(a)G(s)=0$$

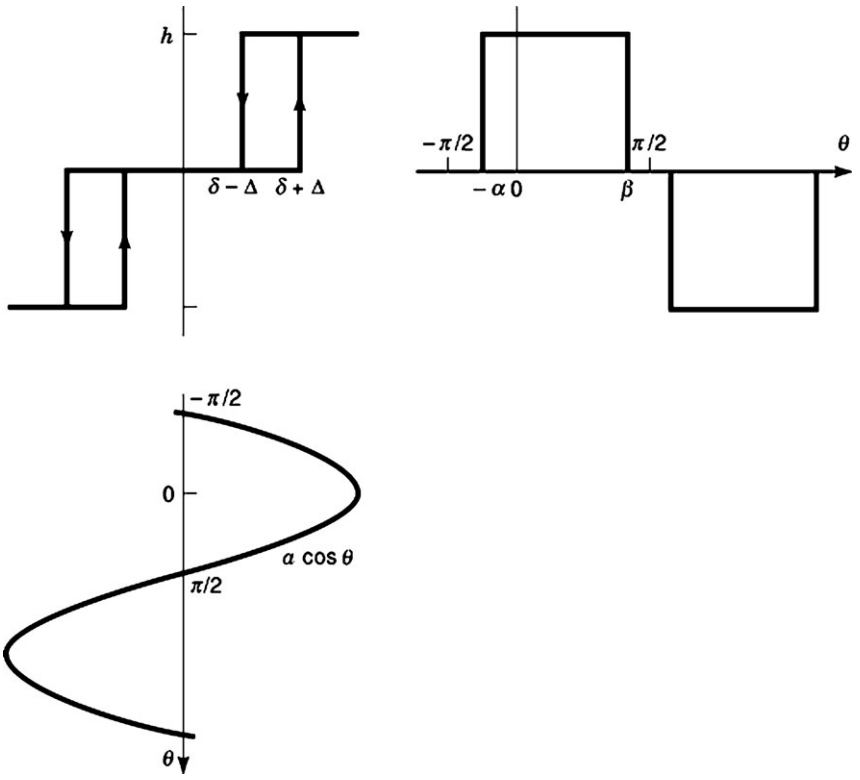

Figure 10. Relay and input/output waveforms.

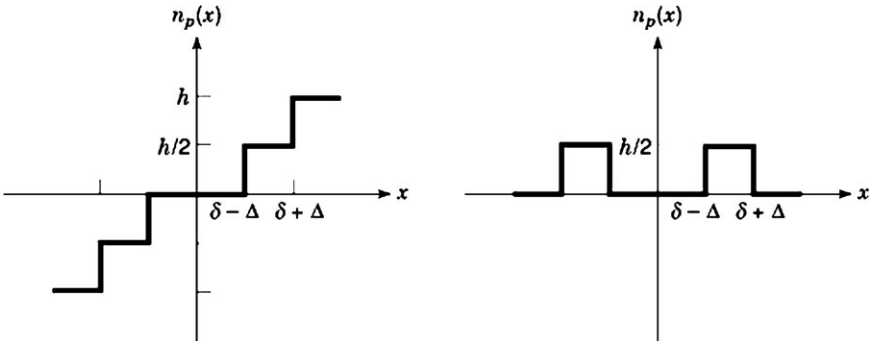

Figure 11. In-phase and quadrature characteristics for the relay of Fig. 10.

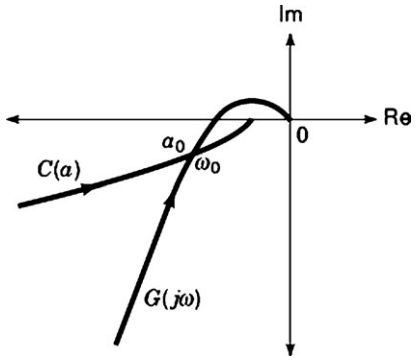

Figure 12. Illustration of limit cycle evaluation.

where $N_{i\gamma}(a)$ is known as the incremental describing function (*IDF*). For a single-valued nonlinearity, $N_{i\gamma}(a)$ can be evaluated from

$$N_{i\gamma}(a) = \int_{-a}^{a} n'(x)p(x)\,dx$$

where $n'(x)$ and $p(x)$ are as previously defined. It can also be shown that $N_{i\gamma}(a)$ is related to $N(a)$ by the equation

$$N_{i\gamma}(a) = N(a) + (a/2)\,dN(a)/da \tag{14}$$

Thus, for example, for an ideal relay, putting $\delta = \Delta = 0$ in Eq. (12) gives $N(a) = 4h/a\pi$, and substituting this value in Eq. (14) yields $N_{i\gamma}(a) = 2h/a\pi$.

As an example of using the DF to investigate the possibility of a limit cycle, consider Fig. 1 with $n(x) = x - (x^3/6)$, $G_c(s) = 1$ and $G_2(s) = K(1 - s)/s(s + 1)$. For this nonlinearity, $N(a) = 1 - (a^2/8)$, so that the $C(a)$ locus starts at -1 on the Nyquist diagram and, as a increases, moves along the negative real axis to $-\infty$ for $a = 2\sqrt{2}$ then, for a greater than this value, the $C(a)$ locus returns along the positive real axis from ∞ to the origin as a becomes large. An os-

cillation will only start to build up, assuming zero initial conditions, if the feedback loop with $G(s)$ alone is unstable since $N(a) \to 1$ for $a \to 0$. This requires the characteristic equation

$$s^2 + s + K - Ks = 0$$

to have a root with a positive real part; that is, $K > 1$. $G(j\omega)$ has 180° phase shift when $\omega = 1$ when its gain is K. Thus the DF solution for the amplitude of the limit cycle is given by

$$|G(j\omega)|_{\omega=1} = \frac{1}{1 - (a^2/8)}$$

which results in

$$K = 8/(8 - a^2)$$

giving

$$a = 2\sqrt{2}[(K-1)/K]^{1/2}$$

As K is increased, because of the shape of the nonlinearity, the limit cycle becomes more distorted. For example, if $K = 2.4$, the DF solution gives $\omega = 1$ and $a = 2.10$, whereas if four harmonics are balanced, which requires a computer program, the limit cycle frequency is 0.719 and the amplitudes of the fundamental, third, fifth, and seventh harmonics at the input to the nonlinearity are 2.515, 0.467, 0.161, and 0.065, respectively.

As the DF approach is a method for evaluating limit cycles, it is sometimes suggested that it cannot be used to guarantee stability of a feedback system, since instability may be exhibited by a signal in the system becoming unbounded, not oscillatory. It is, however, known for the autonomous feedback system of Fig. 1 that if the symmetric odd, single-valued nonlinearity $n(x)$ is sector-bounded such that $k_1 x < n(x) < k_2 x$ for $x > 0$ and $n(x)$ tends to $k_3 x$ for large x, where $k_1 < k_3 < k_2$, then the nonlinear system is either stable or possesses a limit cycle, provided that the linear system with gain K replacing N is stable for $k_1 < K < k_2$. Thus for this situation, which is often true in practice, the nonexistence of a limit cycle guarantees stability.

Accuracy of the Describing Function

Since the DF method is an approximate analytical approach, it is desirable to have some idea of its accuracy. Unfortunate consequences may result if a system is predicted to be stable, and in practice this turns out not to be the case. Although many attempts have been made to find solutions for this problem, those that have been obtained either are difficult to apply or produce results which are often as conservative as the absolute stability criteria discussed earlier.

Since, as has already been shown, the $C(a)$ locus of a sector bounded, single-valued nonlinearity is the diameter of the circle in the circle criterion, errors in the DF method are related to its inability to predict a phase shift which the fundamental may experience in passing through the nonlinearity, rather than an incorrect magnitude of the gain. When the input to a single-valued nonlinearity is a sinusoid together with some of its harmonics, it is easy to show that the fundamental output is not necessarily in phase with the fundamental input; that is, the fundamental gain has a phase shift. The actual phase shift which occurs varies with the harmonic content of the input signal in a complex manner, since the phase shift depends on the amplitudes and phases of the individual harmonic input components.

From an engineering viewpoint, one can therefore obtain a good idea of the accuracy of a DF result by estimating the distortion, d, in the waveform at the input to the nonlinearity. This is relatively straightforward when a limit-cycle solution is obtained since the sinusoidal signal corresponding to the DF solution can be taken as the nonlinearity input and the harmonic content of the signal fed back to the nonlinearity input calculated. Experience indicates that the percentage accuracy of the DF method in predicting the fundamental amplitude and frequency of the limit cycle is usually better than the percentage distortion in the fedback signal.

It is also important to note that the amplitude predicted by the DF is an approximation to the fundamental of the limit cycle, not its peak amplitude. It is possible to estimate the limit cycle more accurately by balancing additional harmonics, as mentioned earlier. Although algebraically this is difficult apart from loops with a nonlinearity having a simple mathematical description—for example, a cubic—it can be done computationally. The procedure involves solving sets of nonlinear algebraic equations, but good starting guesses can usually be obtained for the magnitudes and phases of the other harmonic components from the waveform fedback to the nonlinearity, assuming that its input is the DF solution.

Further Aspects

Before concluding this section on the DF method, it is important to mention two other facets of its application. In introducing the DF, it was indicated that the existence of a limit cycle is usually undesirable; thus if the DF indicates such behavior, the system must be compensated to remove the limit cycle. If the parameters of $n(x)$ and $G_1(s)$, with $G_c(s) = 1$, in Fig. 1 are such that a limit cycle is indicated, because the loci $G_1(j\omega)$ and $C(a)$ intersect, then a compensator $G_c(s)$ can be added with a transfer function such that the loci $G_c(j\omega)G_1(j\omega)$ and $C(a)$ do not intersect. Shaping frequency responses to achieve a specific form is a familiar approach in linear control theory, so this approach can be easily applied. Other approaches such as adding additional feedback paths to compensate for the effect of the nonlinearity may also be possible. This procedure has the advantage, as can the approach of designing a nonlinear controller, of producing an approximately linear system.

A feature of nonlinear systems, as mentioned earlier, is that they possess unique forms of behavior. One such interesting feature is the jump resonance which can occur when a nonlinear feedback system, such as Fig. 1, has a sinusoidal input. Equations can be set up using the DF approach for the feedback loop to balance the harmonic at the frequency of the input sinusoid. Two nonlinear algebraic equations are obtained; and for some situations they can have three, rather than one, solutions for a small

range of input frequencies. The DF can also be used to show that only two of the solutions will be stable, which means that the approximately sinusoidal output from the feedback loop may have two possible values, within this frequency range, which, if it exists, is found near to the resonant frequency of the linearized system. When the input frequency is changed so that the solution of the equations moves from the two-stable-solution to the one-solution (or vice versa) region, a discontinuous change, or jump, in the magnitude of the output may occur.

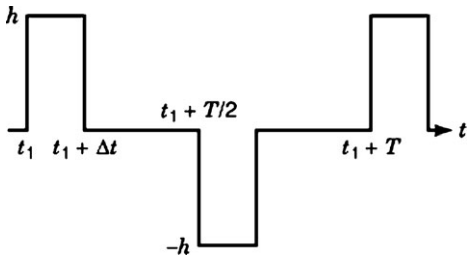

Figure 13. Relay output waveform.

LIMIT CYCLES IN RELAY SYSTEMS

In this section an exact method for the evalution of limit cycles and their stability is discussed which makes use of the fact that the output from a relay is not continuously affected by its input (6, 9). The input only controls the switching instants of the relay and has no further effect on the output until it causes another switching. Therefore to investigate limit cycles in relay systems the analysis starts by assuming a typical relay output waveform, $y(t)$, which for a symmetrical odd limit cycle in a loop having a relay with dead zone and hysteresis takes the form shown in Fig. 13, where T and Δt are unknown and the initial switching is at time t_1. Then to find a possible limit cycle in the autonomous system of Fig. 1 the steady-state response of $G(s)$ to this waveform has to be determined. Several slightly different approaches are possible, but here we follow that used by Tsypkin, primarily because for a relay with no dead zone it allows a simple comparison with the DF method. $y(t)$ is expanded in a Fourier series which gives

$$\begin{aligned} y(t) \quad = \quad & \frac{2h}{\pi}\sum_{n=1(2)}^{\infty}\frac{1}{n}\{\sin(n\omega\Delta t)\cos[n\omega(t-t_1)] \\ & +[1-\cos(n\omega\Delta t)]\sin[n\omega(t-t_1)]\} \end{aligned}$$

The output $c(t)$ is then given by

$$\begin{aligned} c(t) = \frac{2h}{\pi}\sum_{n=1(2)}^{\infty}\frac{g_n}{n}\{&\sin(n\omega\Delta t)\cos[n\omega(t-t_1)+\phi_n] \\ &+[1-\cos(n\omega\Delta t)]\sin[n\omega(t-t_1)+\phi_n]\} \end{aligned} \tag{15}$$

where $g_n = |G(j\omega n)|$ and $\phi_n = \angle G(j\omega n)$. Using the A loci, defined by

$$\begin{aligned} A_G(0,\omega) &= \mathrm{Re}\,A_G(\theta,\omega) + j\mathrm{Im}\,A_G(\theta,\omega) \\ \mathrm{Re}\,A_G(\theta,\omega) &= \sum_{n=1(2)}^{\infty} V_G(n\omega)\sin(n\theta) + U_G(n\omega)\cos(n\theta) \end{aligned} \tag{16}$$

$$\mathrm{Im}\,A_G(\theta,\omega) = \sum_{n=1(2)}^{\infty}\frac{1}{n}\{V_G(n\omega)\cos(n\theta) - U_G(n\omega)\sin(n\theta)\} \tag{17}$$

where $U_G(n\omega) = g_n \cos\phi_n$ and $V_G(n\omega) = g_n \sin\phi_n$. Eq. (15) can be written as

$$c(t) = \frac{2h}{\pi}\{\mathrm{Im}\,A_G(-\omega t + \omega t_1, \omega) - \mathrm{Im}\,A_G(-\omega t + \omega t_1 + \omega\Delta t, \omega)\}$$

Similarly, $c^2(t)$ can be shown to be given by

$$\dot{c}(t) = \frac{2\omega h}{\pi}\{\mathrm{Re}\,A_G(-\omega t + \omega t_1, \omega) - \mathrm{Re}\,A_G(-\omega t + \omega t_1 + \omega\Delta t, \omega)\}$$

To satisfy the above-mentioned switching conditions at times t_1 and $t_1 + \Delta t$, assuming t_1 to be zero without loss of generality, and bearing in mind that $x^2(t_1)$ should be positive and $x^2(t_1 + \Delta t)$negative, we require that

$$\begin{aligned} A_G(0,\omega) - A_G(\omega\Delta t,\omega) \quad &\text{must have} \quad \mathrm{IP} = -\pi(\delta+\Delta)/2h, \\ &\qquad\qquad \mathrm{RP} < 0 \\ A_G(0,\omega) - A_G(-\omega\Delta t,\omega) \quad &\text{must have} \quad \mathrm{IP} = -\pi(\delta-\Delta)/2h, \\ &\qquad\qquad \mathrm{RP} < 0 \end{aligned}$$

where RP and IP denote the real and imaginary parts, respectively. The IP expressions give two nonlinear algebraic equations which, if they have solutions, yield the unknown parameters Δt and T of possible limit cycles. Using these solution values, the corresponding relay input waveform $x(t)$ can be found, from which the RP conditions can be checked, as can the continuity conditions

$$x(t) > \delta - \Delta \quad \text{for } 0 < t < \Delta t \quad \text{and} \quad -(\delta+\Delta) < x(t) < (\delta+\Delta) \quad \text{for } \Delta t < t < T/2$$

to confirm that the relay input signal does not produce switchings other than those assumed.

Since closed-form expressions exist for the A loci of simple transfer functions, analytical solutions can be obtained for the exact frequencies, $1/T$, of limit cycles for some specific systems, especially those in which the relay has no dead zone. Then $\omega\Delta t = \pi$ and the above two nonlinear algebraic equations are identical since only one unknown, T, remains. When the nonlinear algebraic equations are solved computationally the closed-form expressions for the A loci may be used, or their value may be determined by taking a finite number of terms in the series of Eqs. 16 and 17 (6, 9). Another interesting feature of this method is that it is also possible to determine whether a solution to the nonlinear algebraic equations corresponds to a stable or an unstable limit cycle (10). The analysis has assumed a symmetrical odd limit cycle but can be extended to situations where this is not the case. More nonlinear algebraic equations have then to be solved to obtain any possible limit cycle solutions. It is also possible to extend the approach to

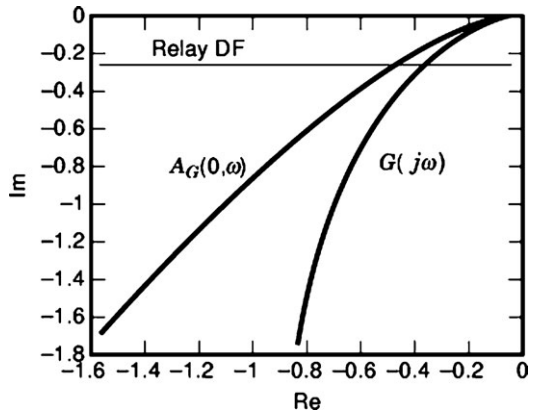

Figure 14. Graphs giving exact and approximate solutions for limit cycle.

find multi-pulse type limit cycles which may exist in relay control of resonant plant transfer functions.

Two Examples

As a simple application of the above method, consider a relay with no dead zone—that is, $\delta = 0$—so that one has the single relationship

$$A_G(0,\omega) \quad \text{must have} \quad \text{IP} = -\pi\Delta/4h, \quad \text{RP} < 0$$

which yields the frequency of the limit cycle. If $G(s) = K/s(1+s\tau)$, then the above expression gives the equation

$$(\pi/2\lambda) - \tanh(\pi/2\lambda) = \Delta/hK\tau$$

where $\lambda = \omega\tau$ for the limit cycle frequency ω. This compares with the DF solution for the same problem, which yields the equation

$$\lambda(1+\lambda^2) = 4hK\tau/\pi\Delta$$

It is also interesting that, since the line with $\text{RP} < 0$ and $\text{IP} = -\pi\Delta/4h$ corresponds to $C(a)$, the negative reciprocal of the DF, the exact and approximate DF solutions can be compared graphically. This is done in Fig. 14, which shows the $G(j\omega)$ and $A_G(0, \omega)$ loci for $K = \tau = 1$ and the $C(a)$ locus for $h/\Delta = 3$. The exact limit-cycle frequency is 1.365 rad/s, and the approximate solution using the DF method is 1.352 rad/s. The accuracy of the DF result may be used to confirm the filter hypothesis, since it can be shown that as τ is increased, thus making $G(s)$ a better low-pass filter, the error in the DF solution for the frequency of the limit cycle decreases.

Consider as a second example a feedback system having a relay with output ± 1 and dead zone ± 1, along with a transfer function $G(s) = 5/s(s^2 + 3s + 1)$. Use of the DF method indicates that the system has two limit cycles, both of frequency 1.000 rads/s, with the larger amplitude one stable and the smaller amplitude one unstable. Two nonlinear algebraic equations need to be solved using the Tsypkin method to find the frequency and pulse width of any limit cycles. Software with graphical facilities is available to do this and the two limit cycles shown in Figs. 15 and Fig. 16 are found. The larger amplitude limit cycle of 15 are found. The larger amplitude limit cycle of 15 is shown

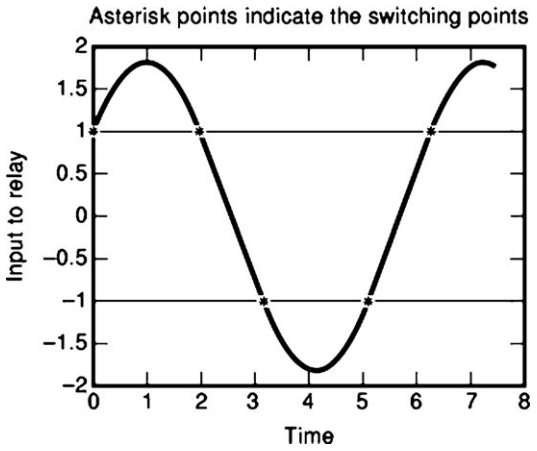

Figure 15. Stable limit cycle solution.

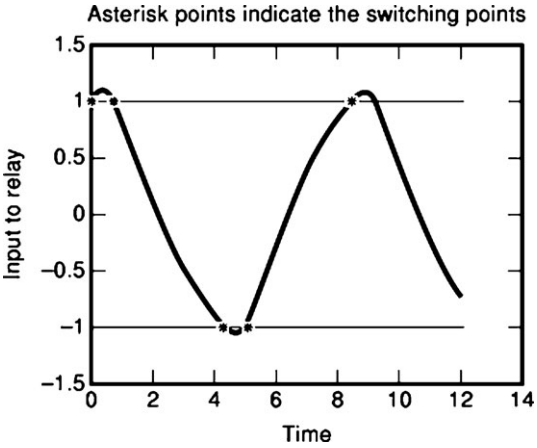

Figure 16. Unstable limit cycle solution.

by the method to be stable with frequency 0.988 rads/s and pulse width 1.967 s, and the smaller amplitude one of Fig. 16 with frequency 0.736 rads/s and pulse width 0.716 s is unstable. It should also be noted that the larger amplitude limit cycle is much closer to a sinusoid so that its frequency is nearer the DF solution of 1.000 rads/s.

SLIDING MODE CONTROL METHODS

For ease of exposition, consider the uncertain linear time invariant system with m inputs given by

$$\dot{x}(t) = Ax(t) + Bu(t) + f(t,x,u) \tag{18}$$

where $A \in \mathrm{R}^{n\times n}$ and $B \in \mathrm{R}^{n\times m}$ with $1 \le m$. Without loss of generality it is assumed that the inputs are independent. The nonlinear function f: $\mathrm{R} \times \mathrm{R}^n \times \mathrm{R}^m \to \mathrm{R}^n$ is assumed to be unknown but bounded by some known functions and represents the parameter uncertainties, nonlinearities, or disturbance signals which are present in the system. Let $\boldsymbol{S}$ be the hyperplane defined by

$$\boldsymbol{S} = \{x \in \Re^n : s(x) = \mathrm{S}x = 0\} \tag{19}$$

where $\boldsymbol{S} \in \mathrm{R}^{m\times n}$ is of full rank. This will define the switching function. It should be noted that the choice of $\boldsymbol{S}$ need

not be restricted to a hyperplane and more general, nonlinear, possibly time-varying switching functions may be chosen.

If there exists a finite time t_s such that the solution to Eq. (61) satisfies

$$s(x) = 0 \qquad \text{for all } t \geq t_s$$

then a sliding motion is taking place for all $t \geq t_s$.

This section will first consider how to design the switching function so that the sliding motion is stable. The problem of designing variable structure control laws so that in finite time the system states are forced on to and subsequently remain on the hyperplane S is considered next. The total insensitivity to a particular class of uncertainty is then demonstrated. The section will conclude with a straightforward example to illustrate the mathematical concepts.

For sliding mode design it is necessary that the system assumes an appropriate canonical form. This so-called regular form is obtained using the orthogonal matrix $T_1 \in \mathrm{R}^{n\times n}$ whereby

$$T_1 B = \begin{bmatrix} 0 \\ B_2 \end{bmatrix}$$

where $B_2 \in \mathrm{R}^{m\times m}$ and is full rank. The transformation matrix exists as B is full rank and can be readily found via QR decomposition. By using the coordinate transformation $x \leftrightarrow T_1 x$ then the nominal linear system can be written as

$$\begin{aligned} \dot{x}_1(t) &= A_{11}x_1(t) + A_{12}x_2(t) \\ \dot{x}_2(t) &= A_{21}x_1(t) + A_{22}x_2(t) + B_2 u(t) \end{aligned} \tag{20}$$

where $x_1 \in \mathrm{R}^{n-m}, x_2 \in \mathrm{R}^m$. Effectively, the system has been decomposed into two connected subsystems, only one of which is directly affected by the system input. If the switching function matrix from Eq. (62) is partitioned compatibly in this coordinate system, then

$$S = [S_1 S_2]$$

where $S_1 \in \mathrm{R}^{m\times(n-m)}$ and $S_2 \in \mathrm{R}^{m\times m}$. During ideal sliding, the motion is given by

$$S_1 x_1(t) + S_2 x_2(t) = 0 \quad \text{for all } t \geq t_s$$

Assuming S_2 is chosen by design to be nonsingular, it follows that

$$x_2(t) = -M x_1(t) \quad \text{for all } t \geq t_s \tag{21}$$

where $M = S^{-1}{}_2 S_1$. It further follows that in the sliding mode, m of the states can be expressed in terms of the remaining $(n-m)$ and thus a reduction in order occurs. The reduced order motion is determined from substituting for $x_2(t)$ from Eq. (21) in Eq. (20) as

$$\dot{x}_1(t) = (A_{11} - A_{12}M)x_1(t) \tag{22}$$

The hyperplane design problem can therefore be considered to be one of choosing a state feedback matrix M to prescribe the required performance to the reduced order subsystem defined by (A_{11}, A_{12}). It can be shown that controllability of the nominal (A, B) pair is sufficient to guarantee controllability of the (A_{11}, A_{12}) pair.

Having defined the switching function, it is necessary to establish sufficient conditions which guarantee that an ideal sliding motion will take place. These will amount to ensuring that in a certain domain enclosing the sliding surface, the trajectories of $s(t)$ must be directed toward it. The associated so-called reachability condition is perhaps most succinctly expressed as

$$s\dot{s} < 0 \tag{23}$$

This choice is readily justified by considering the function

$$V = \frac{1}{2}s^2$$

which is positive definite. Its time derivative along any trajectory is

$$\dot{V} = s\dot{s}$$

It follows that if Eq. (70) holds, then V tends to zero and therefore s tends to zero. This guarantees that a sliding mode is attained. The control signal must thus be defined to satisfy Eq. (70). Subject to this constraint, there are obviously a great many possible control configurations. A common structure is given by

$$u(t) = u_1(t) + u_n(t) \tag{24}$$

where $u_1(t)$ is a linear state feedback law and $u_n(t)$ is a discontinuous or switched component of the form

$$u_n(t) = \rho(t,x)\ \mathrm{sign}(s)$$

The extensive interest in sliding mode control is primarily due to its robustness properties. When sliding, a system is *completely* insensitive to *any* uncertainty which is implicit in the channels of the input distribution matrix; such uncertainty is termed *matched* uncertainty. The reason for this invariance property is easily demonstrated by a consideration of the uncertain state space system

$$\dot{x} = Ax + B(u + f) \tag{25}$$

where f is an unknown but bounded forcing function. In the sliding mode

$$s = sx = 0$$

and thus

$$\dot{s} = s\dot{x} = 0 \tag{26}$$

Substituting from Eq. (75) into Eq. (77),

$$sAx + sB(u_{eq} + f) = 0$$

where u_{eq} is not the applied control signal—which will be of the form of Eq. (73)—but does denote the equivalent linear control that would be required to maintain the sliding mode. This may be expressed as

$$u_{eq} = -(sB)^{-1}sAx - f$$

Substituting this expression for the equivalent control in Eq. (75) yields

$$\dot{x} = (\mathrm{I} - (sB)^{-1}s)Ax$$

The dynamics in the sliding mode are thus completely invariant to the signal f. The system behavior will be determined entirely by Eq. (69) when sliding despite the presence of such matched uncertainty. Any unmatched uncertainty will affect the dynamics of Eq. (69), but such unmatched effects can be minimized by designing M such that the subsystem of Eq. (69) is maximally robust.

In order to illustrate the theoretical concepts of VSCS, consider the double integrator system

$$\ddot{y} = u$$

The system can be expressed in the state-space form

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t)$$

where $x_1 = y$ and $x_2 = y^2$. This system is already in an appropriate regular form for sliding mode design. Consider application of a negative feedback law

$$u = -kx_1(t)$$

The phase portraits when $k = 0.5$ and $k = 1.5$ are shown in Fig. 17. Neither control law yields an acceptable transient when employed as the sole controller; an oscillation is seen to exist in both cases. Consider instead the VSCS defined by

$$u(t) = \begin{cases} -1.5x_1(t) & \text{if } x_1x_2 > 0 \\ -0.5x_1(t) & \text{otherwise} \end{cases}$$

An asymptotically stable motion is seen to result as shown in Fig. 18. By introducing a rule for switching between two control structures, which independently do not provide stability, a stable closed-loop system is formed. Such heuristic arguments can be used to motivate the advantages of a variable structure control approach. However, for design purposes a more logical algorithmic approach is required.<figureAnchor figures="W1024-fig-0017 W1024-fig-0018"/>

Consider now the switching function defined by

$$s = mx_1(t) + x_2(t), \qquad m > 0$$

This is seen to provide a first-order motion in the sliding mode where the pole defining the transient response is determined by the selection of m. The control signal is defined by solving for u from the relationship

$$\dot{s} = -k \operatorname{sgn}(s) \tag{27}$$

This clearly ensures that the reachability condition [Eq. (23)] is satisfied. Essentially, the switching function is differentiated and the resulting state derivatives are replaced with the original system dynamics. Equation (27) thus yields an expression for the control signal in terms of the states and the value of the switching function. The resulting controller will find the switching function at least locally attractive. For simulation, the double integrator model is subject to a disturbance signal $-a_1 \sin(x_1(t))$ which acts in the range of the input distribution matrix. In this way a controller designed for a nominal double integrator model is implemented upon a normalized pendulum system. The design parameters m and k are both set equal to unity. Figure 19 shows the resulting phase plane plot. The system enters a sliding mode and the normalized pendulum is forced to behave as the free first-order system

$$\dot{x}_1(t) = -x_1(t)$$

during this phase of motion. The dynamics in the sliding mode have been wholly specified by the choice of switching function despite the presence of a matched uncertainty contribution.

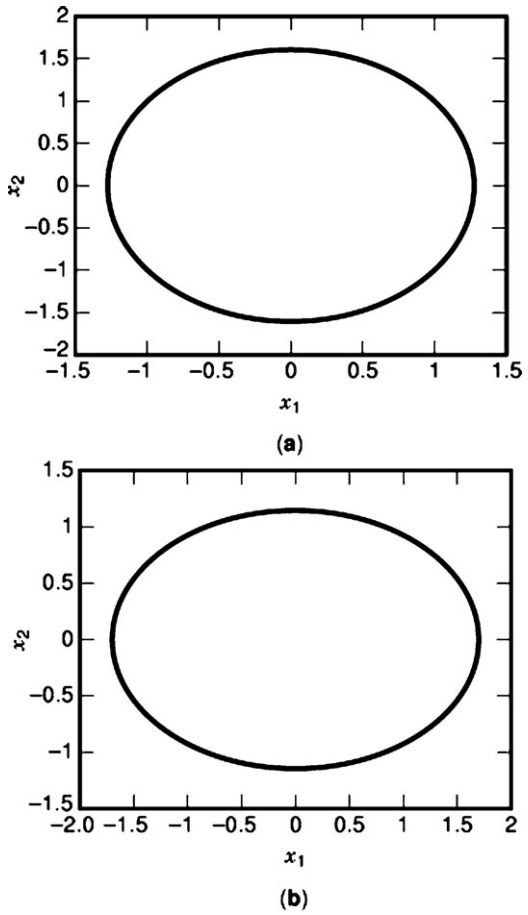

Figure 17. (a) $u = -1.5x_1$; (b) $u = -0.5x_1$. Double integrator control.

OTHER DESCRIBING FUNCTIONS

In a previous section the discussion on describing functions was primarily restricted to the sinusoidal describing function, SDF, since this is used extensively in looking at the effects of nonlinearity in practical systems. Many control systems, however, are subject to inputs or disturbances which cannot be defined deterministically but only as random signals with a given frequency spectrum and amplitude probability density function. The most common amplitude probability density function, $p(x)$, considered is the Gaussian density function which, when it has a zero

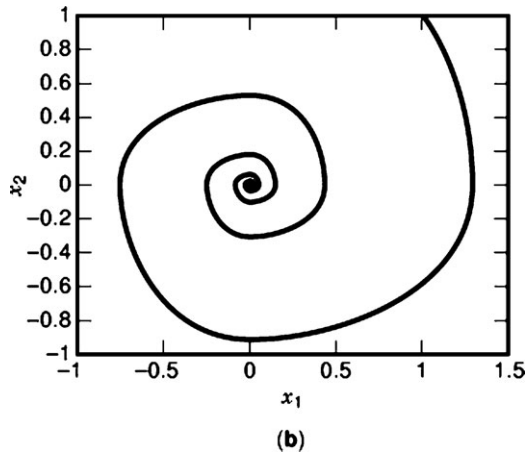

Figure 18. Double integrator with variable structure control system.

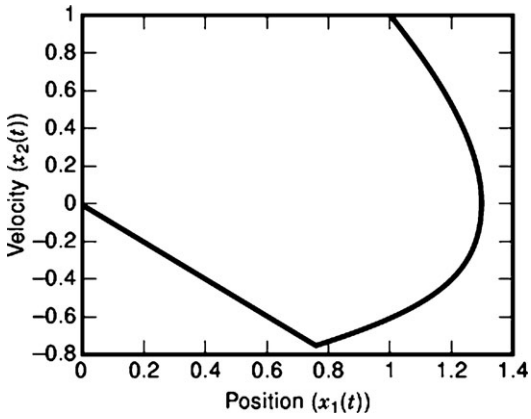

Figure 19. Normalized pendulum with sliding mode control.

mean, is given by

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/2\sigma^2}$$

A general way to define a DF, as mentioned earlier, is that it is that value of gain, K_{eq}, which, when fed with the same input as the nonlinearity $n(x)$, will give a minimum value for the mean-squared error between the nonlinearity output and the gain output. It is then relatively easy to show that

$$K_{eq} = N(\sigma) = \frac{1}{\sigma^2}\int_{-\infty}^{\infty} xn(x)p(x)\,dx$$

when the input signal x has a Gaussian distribution. When this result was first found, the gain was called the equivalent gain, hence the notation K_{eq}, not the random describing function (*RDF*) by which name it is now usually known. The result for x with any zero-mean amplitude probability density function $p(x)$ is

$$K_{eq} = \int_{-\infty}^{\infty} xn(x)p(x)\,dx \Big/ \int_{-\infty}^{\infty} x^2 p(x)\,dx$$

a formula which can be used for a single-valued nonlinearity with sinusoidal or Gaussian inputs when the appropriate $p(x)$ is used. When the input x is a sinusoidal or Gaussian signal, however, it can also be shown that the error signal between the nonlinearity and linear gain outputs—that is, $n(x) - K_{eq}x$—is uncorrelated with the input x (6). Typically, when dealing with Gaussian inputs to a simple nonlinear feedback system, the mean-squared values of the signals at various points in the loop can be calculated approximately using the RDF for the nonlinearity.

In many feedback systems it may be necessary to take account of bias, as well as other signals, due to constant input or disturbance signals or because of asymmetry in nonlinear characteristics. In this case the nonlinearity, again using the minimum mean-squared error definition, may be modeled by two DFs, one for the bias, γ, and one for the other input (6, 7). When the other input is considered as a sinusoid of amplitude a, then the two DFs for the single-valued nonlinearity $n(x)$ are given by

$$N(a,\gamma) = (2/a^2)\int_{-a}^{a} xn(x+\gamma)p(x)dx$$

and

$$N_\gamma(a,\gamma) = (1/\gamma)\int_{-a}^{a} n(x+\gamma)p(x)dx$$

the former being the DF for the sinusoid and the latter for the bias. Here $p(x) = 1/\pi(a^2 - x^2)^{1/2}$ for the sinusoidal signal. Use of this DF allows, amongst other possibilities, for the determination of limit cycles with bias. For example, if in Fig. 1 the input $r(t)$ has a constant value R, then balancing the bias and fundamental of the limit cycle gives the two equations

$$RG_c(0) - \gamma N_\gamma(a,\gamma)G(0) = \gamma$$
$$1 + N(a,\gamma)G(j\omega) = 0$$

The equations can be solved to find the bias γ and sinusoidal amplitude a of the limit cycle at the input to the nonlinearity.

The above approach can be used in principle to obtain a DF representation for a nonlinearity whose input consists of any number of uncorrelated signals, but for practical reasons the approach is difficult to justify for more than two or possibly three components. A difficulty in applying such multiple input describing functions is understanding the errors which are caused by neglecting not only "higher harmonics of the input signals" but also "cross-modulation products" which may be produced at frequencies lower than those in the nonlinearity input signal.

Reasonably successful use, however, of the DF approach for two related frequency sinusoidal inputs to a nonlinearity has been achieved to give results of some value in control system design. This requires consideration of two inputs such as a $\cos \omega t$ and $b\cos(3\omega t + \phi)$ so that the describing functions for the two signals become functions of the three parameters a, b and ϕ, not just a and b (6). Analytically, results can only be obtained for simple nonlinearities such as a cubic; but by using computational methods, other characteristics can be considered (11). This procedure has been used to investigate subharmonic oscillations and synchronization phenomena when the feedback loop of Fig. 1 has an input $r(t)$ which is sinusoidal, and it

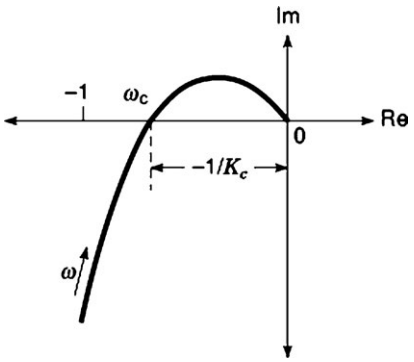

Figure 20. K_c and w_c from the Nyquist plot.

has also been used for the more accurate determination of limit cycle waveforms by balancing both the fundamental and another harmonic, usually the third.

RELAY AUTOTUNING

A relatively recent application of the DF approach in control system design is its use in the relay autotuning method for setting PID controller parameters. The procedure is very useful in those cases where it is difficult to obtain a good mathematical model for the process or retuning has to be done on site by an operator. The basic concept employed is that knowledge of the critical frequency, ω_c, and gain, K_c, of a process illustrated on a plant frequency response shown in Fig. 20 can often provide sufficient information for setting the parameters of a PID controller. The approach was suggested many years ago by Ziegler and Nichols (12), where K_c and ω_c were found by placing the PID controller in the P mode and adjusting the gain until oscillation took place. There were difficulties doing this in practice, however, one problem being that the oscillation amplitude is only limited by saturation in the controller, actuator or plant. More recently therefore, Astrom and Hagglund (13) recommended estimating ω_c and K_c from results obtained by replacing the controller with an ideal relay, a feature, as illustrated in Fig. 21, which can easily be built into a modern microprocessor controller. When this is done the amplitude of the limit cycle, a, at the relay input and its frequency, ω_c, are measured. Then according to DF theory, $\omega_c = \omega_0$ and $K_c = 4h/a\pi$. Strictly speaking, a should be the amplitude of the fundamental frequency component, and the results are only exact when the limit cycle is sinusoidal. In many cases, however, these formulae provide reasonable estimates for ω_c and K_c which may then be used in an appropriate algorithm (many of which have been put forward recently) for setting the PID parameters. If the form of the plant transfer function is known but not its parameters, then it may be possible, certainly for low-order transfer functions, to make use of the known Tsypkin solution for the limit cycle to estimate the plant parameters. When the plant has several unknown parameters, more than one autotuning test may have to be done using different values of hysteresis in the relay or possibly with bias signals introduced into the loop.

Obviously, the relay on–off levels control the limit cycle amplitude; and if these are varied, some information may be found about any nonlinearity in the plant (14, 15). In such cases it may be possible to make the system behavior more linear by incorporating appropriate nonlinearity in the PID elements of the controller.

MULTIVARIABLE SYSTEMS

So far in this presentation, only simple nonlinear feedback systems such as Fig. 1 have been considered, apart from when discussing the phase plane. In principle there is no difficulty in extending use of both the DF and Tsypkin approaches to feedback loops with more than one nonlinear element, although for the latter approach the nonlinearities must be of the relay type. The problem with using the sinusoidal describing function approach is that the assumption of a sinusoidal input to all the nonlinearities must be reasonable for the situation under investigation. For some configurations of nonlinearities and linear dynamic elements in a feedback loop, this will not be true. Using the Tsypkin approach, more nonlinear algebraic equations are formulated and their possible solutions must be investigated (6, 16).

Several investigators have produced results for studies on the multivariable, typically two-input–two-output version of Fig. 1. Here the nonlinearity consists of four individual, or in many cases only two on the diagonal, nonlinear elements. Using describing function analysis for this configuration can often be justified since good filtering may exist for all the—or in the case of two input–two output, two—feedback loops. Software written to investigate such systems, using both the DF and Tsypkin methods, has been described in the literature (17–19). Absolute stability results, similar to those given earlier but which result in more complicated graphical procedures, have been extended to the two-input–two-output multivariable system (20). Like the situation for the single-input–single-output system, however, the results they produce are often very conservative and may not be of value for practical problems.

SLIDING MODE CONTROL

The detailed discussion here comments further on uncertain systems. It is obviously desirable to consider other nominal nonlinear system descriptions in order to broaden the applicability of the method. Some work in this area has considered problems relating to a particular application area—for example, robotics—and has developed sliding mode controllers for such specific classes of nonlinear system (21). Other work has tried to consider more general nonlinear system descriptions. Some of the conditions placed upon the particular system representation can be quite restrictive. One example of this is the class of feedback linearizable systems. It is possible to augment the traditional linearizing feedback with a sliding mode component which will provide robustness to some uncertainty in the sliding mode. However, the conditions which must be satisfied to feedback linearize the system initially are quite restrictive and so limit the applicability of the methodol-

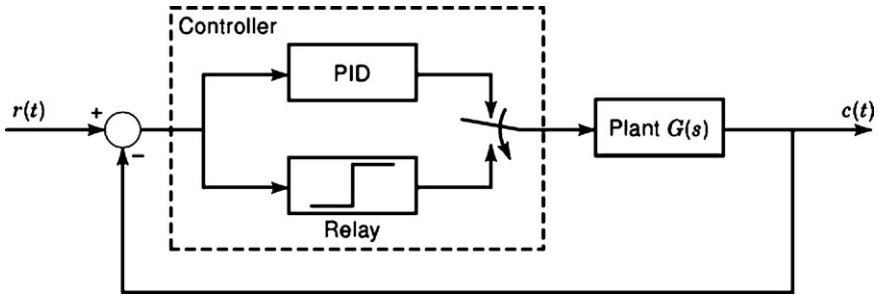

Figure 21. Block diagram for relay autotuning.

ogy. Perhaps the most widely applicable method to date has resulted from the development of sliding mode control schemes for differential input–output system representations (22). These yield dynamic controllers which act as a natural filter on any discontinuous elements of the control signal and are applicable to a fairly broad class of nonlinear systems. This is thus a considerable breakthrough in the development of robust controllers for nonlinear system descriptions.

The previous exposition relates to state-feedback based sliding mode control schemes (23). For practical application, controllers based upon measured output information are required (24). There are two ways to approach this problem. A dynamical system, or observer, may be used to estimate unmeasurable system states. Because of the inherent robustness of sliding mode controllers, some significant work has considered the development of an associated sliding mode observer (25, 26). The robustness properties have been shown to transfer. However, there are restrictions on the (A, B, C) triple used for observer design. In particular, the invariant zeros of (A, B, C) are shown to play a crucial role in determining the zero dynamics in the sliding mode. It thus follows that these invariant zeros must be stable. Despite this restriction, the closed-loop robustness properties of such a sliding mode controller–observer scheme are excellent. The sliding observer is also finding a promising area of application relating to fault detection and isolation. The second approach to output feedback based sliding mode control is to develop design methodologies which produce output dependent control strategies (27). This restricts the class of systems which may be considered as the switching surface must be output-dependent, and thus S must lie in the range of the output distribution matrix. Again the sliding mode dynamics will be dependent upon the location of the system transmission zeros. The development of design methodologies and associated case studies relating to output feedback-based sliding mode control strategies and sliding mode controller–observer strategies require further development. The development of sliding mode controllers based upon output measurements for nonlinear systems is very much an open research problem.

BIBLIOGRAPHY

1. K. S. Narendra, J. H. Taylor, *Frequency Domain Criteria for Absolute Stability*, New York: Academic Press, 1973.
2. V. M. Popov, Absolute stability of nonlinear control systems of automatic control, *Autom. Remote Control*, **22**: 857–858, 1962.
3. I. W. Sandberg, A frequency domain condition for the stability of feedback systems containing a single time-varying nonlinear element, *Bell Syst. Tech. J.*, **43** (4): 1601–1608, 1964.
4. D. P. Atherton, D. H. Owens, Boundedness properties of nonlinear multivariable feedback systems, *Electron. Lett.*, **15** (18): 559ff, 1979.
5. Y. S. Cho, K. S. Narendra, An off-axis circle criterion for the stability of feedback systems with a monotonic linearity, *IEEE Trans. Autom. Control*, **13** (4): 413–416, 1968.
6. D. P. Atherton, *Nonlinear Control Engineering: Describing Function Analysis and Design*, London: Van Nostrand-Reinhold, 1975.
7. A. Gelb, W. E. Vander Velde, *Multiple-Input Describing Functions and Nonlinear System Design*, New York: McGraw-Hill, 1996.
8. S. K. Choudhury, D. P. Atherton, Limit cycles in high-order nonlinear systems, *Proc. Inst. Electr. Eng.*, **121**: 717–724, 1974.
9. P. A. Cook, *Nonlinear Dynamical Systems*, London: Prentice-Hall International, 1986.
10. R. Balasubramanian, Stability of limit cycles in feedback systems containing a relay, *IEE Proc., Part D*, **1**: 24–29, 1981.
11. J. C. West, J. L. Douce, R. K. Livesley, The dual input describing function and its use in the analysis of nonlinear feedback systems, *Proc. Inst. Electr. Eng., Part B*, **103**: 463–474, 1956.
12. J. G. Ziegler, N. B. Nichols, Optimum settings for automatic controllers, *Trans. ASME*, **64**: 759–768, 1942.
13. K. J. Astrom, T. Hagglund, Automatic tuning of simple regulators with specifications on phase and amplitude margins, *Automatica*, **20** (5): 645–651, 1984.
14. J. H. Taylor, K. J. Astrom, *A Nonlinear PID Autotuning Algorithm*, Seattle, WA: ACC, 1986, pp. 1–6.
15. D. P. Atherton, M. Benouarets, O. Nanka-Bruce, Design of nonlinear PID controllers for nonlinear plants, *Proc. IFAC World Cong. '93*, Sydney, Australia, 1993, Vol. 3, pp. 355–358.
16. D. P. Atherton, Conditions for periodicity in control systems containing several relays, *3rd IFAC Cong.*, London, 1966, Paper 28E.
17. J. H. Taylor, Applications of a general limit cycle analysis method for multivariable systems, inR. V. Ramnath, J. K. Hedrick, andH. M. Paynter (eds.), *Nonlinear System Analysis and Synthesis*, New York: ASME, 1980, Vol.**2**.
18. D. P. Atherton *et al.*, Suns, the Sussex University Nonlinear Control Systems Software, *3rd IFAC/IFAD Symp. Comput. Aided Des. Control Eng. Syst.*, Copenhagen, 1985, pp. 173–178.
19. O. P. McNamara, D. P. Atherton, Limit cycle prediction in free structured nonlinear systems, *IFAC Cong.*, Munich, 1987, Vol. 8, pp. 23–28.
20. D. P. Atherton, *Stability of Nonlinear Systems*, New York: Wiley, Research Studies Press, 1981.

21. J. J. E. Slotine, W. Li, *Applied Nonlinear Control*, London: Prentice-Hall International, 1991.
22. H. Sira Ramirez, On the dynamical sliding mode control of nonlinear systems, *Int. J. Control*, **57**: 1039–1061, 1993.
23. V. I. Utkin, *Sliding Modes in Control Optimisation*, Berlin: Springer-Verlag, 1992.
24. C. Edwards, S. Spurgeon, *Sliding Mode Control: Theory and Applications*, London: Taylor & Francis, 1997.
25. J. J. E. Slotine, J. K. Hedrick, E. A. Misawa, On sliding observers for nonlinear systems, *J. Dyn. Syst. Meas. Control*, **109**: 245–252, 1987.
26. C. Edwards, S. K. Spurgeon, On the development of discontinuous observers, *Int. J. Control*, **59**: 1211–1229, 1994.
27. S. Hui, S. H. Zak, Robust output feedback stabilisation of uncertain systems with bounded controllers, *Int. J. Robust Nonlinear Control*, **3**: 115–132, 1993.

DEREK ATHERTON
SARAH SPURGEON
School of Engineering
University of Sussex,, Brighton,
England, BN1 9QT
Department of Engineering
University of Leicester,,
Brighton, England, BN1 9QT