Pilot Project No: CZ/02/B/F/PP/134001

Course on Dynamics of multidisplicinary and controlled Systems

Part IV
Nonlinear Control Systems

October 26, 2004

Derek P. Atherton

Department of Engineering and Design

School of Science and Technology

University of Sussex

Contents

Module 1

Introduction to Nonlinear systems.

Module units

Module overview.

This module introduces nonlinearity and the types of nonlinearity which occur in control systems. It discusses some possible effects nonlinearity might have on open loop and closed loop system behavior.

Module objectives. *When you have completed this module you should be able to:*

1. *Understand nonlinearity and how it might occur in dynamic systems*
2. *Understand the possible effects nonlinearity might have on the dynamic behavior of a control system*

Module prerequisites. *First year mathematics, Fourier series, transfer functions, frequency response*

1.1 Introduction

Extensive theoretical techniques for the analysis and design of linear control systems have been developed over the last 50 years. Unfortunately in practice all systems exhibit nonlinear behavior and the use of linear analysis only, may not provide an adequate description of the behavior. Linear systems have the important property that they satisfy the superposition principle. This leads to many important advantages in methods for their analysis. For example, in a simple feedback loop with both set point and disturbance inputs their affect on the output when they are applied simultaneously is the same as the sum of their individual effects when applied separately. This would not be the case if the system were nonlinear. Thus, mathematically a linear system may be defined as one which with input $x(t)$ and output $y(t)$ satisfies the property that the output for an input $ax_1(t)+bx_2(t)$ is $ay_1(t)+by_2(t)$, if $y_1(t)$ and $y_2(t)$ are the outputs in response to the inputs $x_1(t)$ and $x_2(t)$, respectively, and a and b are constants. A nonlinear system is defined as one which does not satisfy the superposition property. The simplest form of nonlinear system is the static nonlinearity where the output depends only on the current value of input but in a nonlinear manner, for example the mathematical relationship

$$y(t) = ax(t) + bx^3(t) \tag{1.1}$$

where the output is a linear plus cubed function of the input.

More commonly the relationship could involve both nonlinearity and dynamics so that it might be described by the nonlinear differential equation

$$\frac{d^2y}{\mathrm{d}t^2} + a\left[\frac{dy}{dt}\right]^3 + by(t) = x(t) \tag{1.2}$$

From an engineering viewpoint it may be desirable to think of this equation in terms of a block diagram consisting of linear dynamic elements and a static nonlinearity, which in this case is a cubic with input dy/dt and output $a\,(dy/dt)^3$. A major point about nonlinear systems, however, is that their response is amplitude dependent so that if a particular form of response, or some measure of it, occurs for one input magnitude it may not result for some other input magnitude. This means that in a feedback control system with a nonlinear plant if the controller designed does not produce a linear system then, to adequately describe the system behavior one needs to investigate the total allowable range of the system variables. For a linear system one can claim that a system has an optimum response, assuming optimum is precisely defined, for example minimization of the integral squared error, using results obtained for one input amplitude. On the other hand for a nonlinear system the response to all input amplitudes must be investigated and the optimum choice of parameters to minimize the criterion will be amplitude dependent. Perhaps the most interesting aspect of nonlinear systems is that they exhibit forms of behavior not possible in linear systems and more will be said on this later.

1.2 Forms of nonlinearity

Nonlinearity may be inherent in the dynamics of the plant to be controlled or in the components used to implement the control. For example, there will be a limit to the torque obtainable from an electric motor or the current that may be input to an electrical heater, and indeed good design will have circuits to ensure this is the case to avoid destruction of a component. Sizing of components must take account of both the required performance and cost, so it is not unusual to find that a rotary position control system will develop maximum motor torque for a demanded step angle change of only a few degrees. Also, although component manufacture has improved greatly in the last decades flow control valves possess a dead zone due to the effect of friction and there characteristics on opening and closing are not identical due to the flow pressure. Improved design might produce more linear components but at greater cost so that such strategies would not be justified economically. Alternatively one may have nonlinear elements intentionally introduced into a design in order to improve the system specifications, either from a technical or economic viewpoint. A good example is the use of relay switching. Identifying the precise form of nonlinearity in a system component may not be easy and like all modeling exercises the golden rule is to be aware of the approximations in a nonlinear model and the conditions for its validity.

Friction always occurs in mechanical systems and is very difficult to model, with many quite sophisticated models having been presented in the literature. The simplest is to assume the three components of stiction, an abbreviation for static friction, Coulomb friction and viscous friction. As its name implies stiction is assumed to exist only at zero differential speed between the two contact surfaces. Coulomb friction with a value less than stiction is assumed to be constant at all speeds, and viscous friction is a linear effect being directly proportional to speed. In practice there is often a term proportional to a higher power of speed, and this is also the situation for many rotating shaft loads, for example a fan. Many mechanical loads are driven through gearing rather than directly. Although geared drives, like all areas of technology, have improved through the years they always have some small backlash. This may be avoided by using anti-backlash gears but these are normally only employed for low torques. Backlash is a very complicated phenomenon involving impacts between surfaces and is often modeled in a very simplistic manner. For example, a simple approach often used consists of an input- output position characteristic of two parallel straight lines with possible horizontal movement between them. This makes two major assumptions, first that the load shaft friction is high enough for contact to be maintained with the drive side of the backlash when the drive slows down to rest. Secondly when the drive reverses the backlash is crossed and the new drive side of the gear 'picks up' the load instantaneously with no loss of energy in the impact and both then move at the drive shaft speed. Clearly both these assumptions are never true in practice so their limitations need to be understood and borne in mind when using such an approximation.

The most widely used intentional nonlinearity is the relay. The on-off type, which can be described mathematically by the signum function, that is it switches on if its input exceeds a given value and off if it goes below the value, is widely used normally with some hysteresis between the switching levels. Use of this approach provides a control strategy where the controlled variable oscillates about the desired level. The switching mechanism varies significantly according to the application from electromechanical relays at low speed to fast electronic switches employing transistors or thyristors. A common usage of the relay is in temperature control of buildings, where typically the switching is provided by a thermostat having a pool of mercury on a metal expansion coil. As the temperature drops the coil contracts and this

causes a change in angle of the mercury capsule so that eventually the mercury moves, closes a contact and power is switched on. Electronic switching controllers are used in many modern electric motor drive systems, for example, to regulate phase currents in stepping motors and switched reluctance motors and to control currents in vector control drives for induction motors. Relays with a dead zone, that is three position relays giving positive, negative and a zero output are also used. The zero output allows for a steady state position within the dead zone but this affects the resulting steady state control accuracy.

Theoretical analysis and design methods for nonlinear systems typically require a mathematical model for the nonlinearity, sometimes of a specific form, say a power law or polynomial representation. The mathematical model used will be an approximation to the true situation.

Typical models used are:

(i) a simple mathematical function

(ii) a series approximation

(iii) a discontinuous set of functions, typically straight lines to produce a linear segmented characteristic.

1.3 Structure and behaviour

From a control engineering viewpoint there are major benefits to be gained if something is known about the structure of a nonlinear plant compared with treating it as a black box. The relevance of this is perhaps best explained by considering the nonlinearity $x + ax^3$, and a dynamic element with transfer function $K/s(s + b)$, where a and b are constants. If they are placed in cascade and a sinusoidal input applied the output will be a deterministic waveform containing two frequency components one at the same frequency as the input and the other at three times that frequency. Any cascade combination of linear and nonlinear elements will always produce a deterministic output for a periodic input, which in principle can be calculated for any given discrete input frequency spectrum. New frequency components can only be created by the nonlinearities and the linear elements simple alter the relative magnitudes and phases of these components.

For example, if the above nonlinearity is placed before and after a linear transfer function and a sinusoid of frequency, f, is applied at the input, then the input to the second nonlinearity will consist of the fundamental, f, plus third harmonic, $3f$, with magnitude and phase dependent on the frequency. These two frequencies applied to the second nonlinearity will produce an output containing the frequency components, f, $3f$, $5f$, $7f$, and $9f$. One could define a frequency response for such a cascade structure of linear and nonlinear elements as the ratio of the output at the applied fundamental frequency, f, to the input sinusoid at this frequency. The result, as for a linear system, would be a magnitude and phase, which varies with, f, but because of the nonlinearities it would also vary with the amplitude of the sinusoid. Thus an approximate model for the combination could be portrayed graphically by a set of frequency response plots for different input amplitudes, or gain and phase plots against amplitude for different frequencies. For many problems encountered in control engineering this may prove to be a reasonably good model since many dynamic elements have a low pass filtering properties so that the frequency, f, predominates at the output. With no linear dynamic elements in the combination then these latter plots would be the same for all frequencies and the approximate, first harmonic, or quasi-linearized, model would be gain and phase curves as a function of amplitude. This representation of a nonlinear element is known as a describing function and is considered later.

If alternatively the simple feedback loop of Fig. 1.1 is considered with the nonlinearity being the aforementioned $x + ax^3$, the controller $G_c(s) = 1$, and the plant transfer function $G(s) = K/s(s + b)$, then very different situations are possible for the response to a sinusoidal input. Dependent on the values of a, b and K both assumed positive, and the amplitude and frequency of the input, the output may be (a) approximately sinusoidal with the same frequency as the input, similar to the above cascade connection; (b) approximately sinusoidal with a frequency one third of the input, known as a subharmonic oscillation; or (c) a waveform known as chaotic, which is not definable mathematically but completely repeatable for the same initial conditions. These behaviors are unique to nonlinear feedback systems, aspects which make such systems extremely interesting, but in addition it has meant that no general analytical methods are available for predicting their behavior.

Since Fig. 1.1 may be regarded as the structure of many simple nonlinear feedback control loops it is relevant to discuss further it's possible forms of behavior. In doing so it will be assumed that any nonlinearity and any linear transfer function can exist in the appropriate blocks, but the form they must

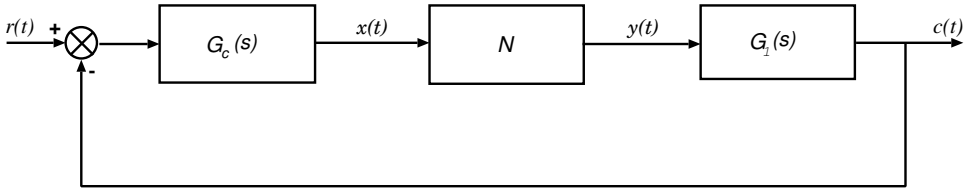

Figure 1.1: Simple nonlinear feedback system

take so that the loop has one of the specific properties mentioned will not be discussed. Firstly, 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. In particular, instead of reaching a stationary equilibrium point, the system may move from some initial conditions into a limit cycle, a continuous oscillation which can be reached from several possible initial conditions. This behavior is distinct from an oscillation in an idealized linear system since the amplitude of this oscillation depends 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, as mentioned earlier, 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 with an amplitude dependent upon the past history of the input or the initial conditions of the system, the so-called jump phenomena in the frequency response. 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, a subharmonic oscillation as mentioned earlier. Also if, for example, the autonomous system had a limit cycle, then the addition of a sinusoidal input could cause the limit cycle frequency to change or 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 exist. 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 operations on the system, typically the error, signals. Many systems can be made to operate satisfactorily with the addition of a simple controller, $G_c(s)$ in Fig. 1.1. 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.

1.4 Analysis Methods

Unlike linear systems there are no completely general methods for the analysis of nonlinear systems. Those to be covered in the following sections are those which most easily link to the topics covered in the earlier modules and are easily applied to, and provide understanding of, practical situations, particularly when combined with simulation studies. The first topic discussed in module 2 is the phase plane method, which can normally only be used to investigate second order systems. It is a useful technique, because it can be used when more than one nonlinearity exists in a system and also because many control problems, such as position control systems, can be modeled approximately by second order dynamics.

The describing function approach is discussed in module 3. It is a non-exact method whereby the

nonlinearity is approximated by a gain, which depends on the amplitude of the input signal, usually taken as a sinusoid. It can, however, enable the designer to obtain insight and understanding of a nonlinear situation and, of course, the ideas can often be further checked by simulation. The describing function approach, since a limit cycle is a form of instability found in nonlinear systems and can be predicted approximately by the describing function method, is useful for stability studies. The method 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 mentioned previously to be investigated. Module 4 discusses two topics, the first being the determination of limit cycles in a feedback loop. The second is an introduction to methods for evaluating the absolute stability of the system shown in Fig. 1.1.

Module 2

The Phase Plane

Module units

Module overview.

This module introduces the basic concepts of the phase plane approach for studying the transient response of second order systems. Particular emphasis is placed on its use for control systems where the nonlinearity may be approximated by a linear segmented characteristic.

Module objectives. *When you have completed this module you should be able to:*

1. Understand phase plane portraits
2. Obtain the transient behavior for some simple systems using the phase plane approach

Module prerequisites. *First year mathematics, differential equations, time response of second order systems*

2.1 Overview of the phase plane method for second order systems.

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

$$\begin{aligned} \dot{x}_1 &= P(x_1, x_2) \\ \dot{x}_2 &= Q(x_1, x_2) \end{aligned} \tag{2.1}$$

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

$$\frac{dx_2}{dx_1} = \frac{\dot{x}_2}{\dot{x}_1} = \frac{Q(x_1, x_2)}{P(x_1, x_2)} \tag{2.2}$$

A second order nonlinear differential equation representing a control system with a 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 $x_1 = x,\ x_2 = \dot{x}$ then it can be written as

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= f(x_1, x_2) \end{aligned} \tag{2.3}$$

which is a special case of eqn. 2.2. A variety of procedures were proposed, before the advent of good computation facilities, for sketching state (phase) plane trajectories for eqns. 2.1 and 2.3. A complete plot showing trajectory motions throughout the entire state (phase) plane from different initial conditions is known as a state (phase) portrait. Phase portraits, or specific responses in the phase plane, are easily obtained with modern simulation facilities and they are often more helpful for understanding the system behavior than displays of the variables x_1 and x_2 against time.

If $f(x_1, x_2) = 2\,\zeta x_2 + x_1$, then one has the normalized second order system with damping ratio ζ and phase plane trajectories are shown in Fig. 2.1 from the initial condition $(-1, 0)$ for $\zeta = 0, 0.5$ and 2.0 respectively. For $\zeta = 0.5$ and 2.0 the origin is a stable singular point, known as a focus for $\zeta = 0.5$ and a node for $\zeta = 2.0$, to which the trajectories converge. For the oscillatory case of no damping the origin is called a center.

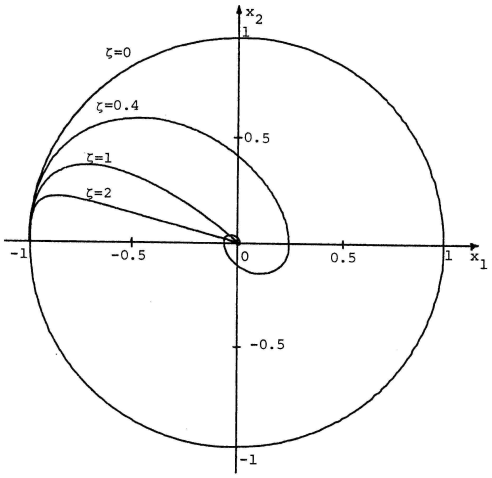

Figure 2.1: Phase trajectories of linear second order system

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

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

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

$$\frac{dx_2}{dx_1} = \frac{\dot{x}_2}{\dot{x}_1} = \frac{\mu(1 - x_1^2)x_2 - x_1}{x_2} \tag{2.4}$$

which is only singular (that is at an equilibrium point), when the right hand side of eqn. 2.4 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 $\mu > 2$. Fig. 2.2 shows a simulation result for $\mu = 1$ from a small initial condition which leads to a stable limit cycle solution.

Many nonlinear effects in control systems, such as saturation, friction etc., are best approximated by linear segmented characteristics rather than continuous mathematical functions. Use of the phase

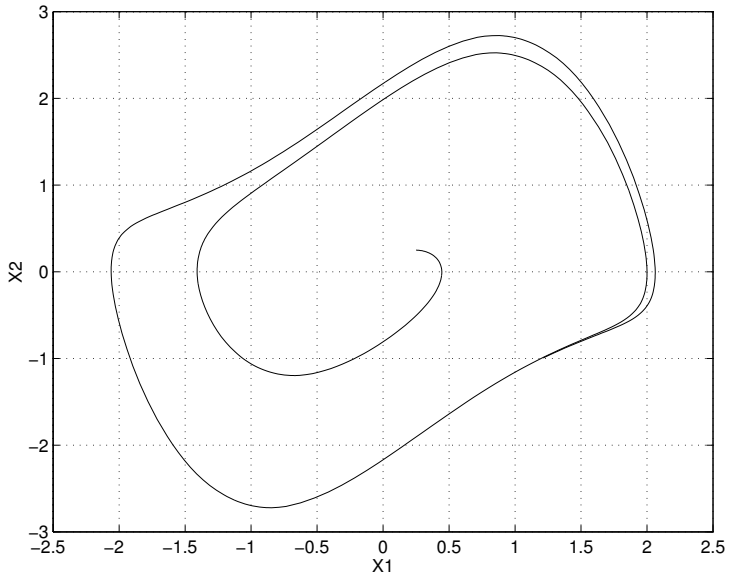

Figure 2.2: Limit cycle for Van der Pol equation

plane to study the effects of this type of nonlinearity on the system response was one of the original contributions to the phase plane method in control. It has the advantage that it results in a phase plane divided up into different regions but with a linear differential equation describing the motion in each region.

2.2 Phase plane methods for systems with linear segmented characteristics

To illustrate the approach for systems which contain nonlinear elements describable by linear segmented characteristics two examples are given below.

Example 1

Consider a simple relay position control system with nonlinear velocity feedback having the block diagram shown in Fig. 2.3.

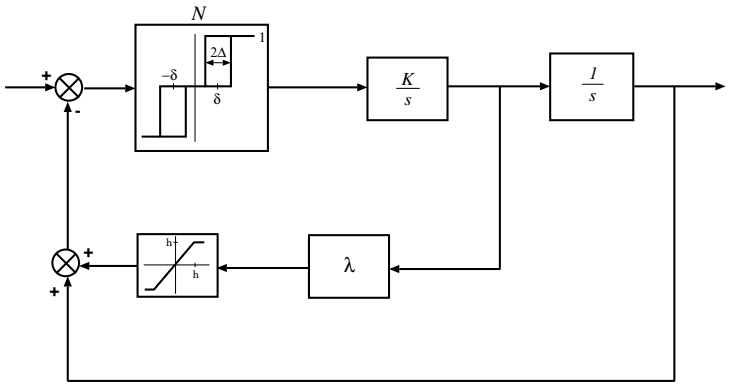

Figure 2.3: Block diagram of relay position control system

Initially it is assumed 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, the velocity $\dot{x}_1$ by x_2, it is seen that the relay output of ± 1 or 0 is equal to $\ddot{x}_1/K$, and that the relay input is equal to $-x_1 - \lambda x_2$. Taking the dead zone of the relay $\pm\delta$ to be equal to ± 1, the

motion of the system is described by

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

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. 2.4 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 co-ordinates x_1 and x_2 is

$$x_2^2 - x_{20}^2 = 2K(x_1 - x_{10}) \tag{2.5}$$

where x_{10} and x_{20} are the initial values of x_1 and x_2. Eqn. 2.5 describes parabolas for K both positive and negative and when the relay output is zero the solution is $x_2 = x_{20}$, which corresponds to a constant velocity motion, since the block after the relay is ideal in neglecting any friction.

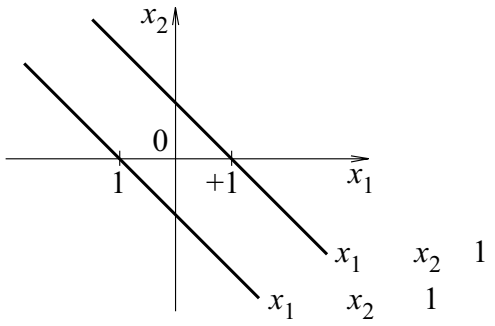

Figure 2.4: Graph showing phase plane boundaries.

If viscous friction were taken into account then the transfer function of this block would be of the form $K/(s+a)$ and the velocity would reduce. For the simple equations of the idealized system it is easy to calculate the system's response from any initial condition (x_{10}, x_{20}) in the phase plane. Fig. 2.5 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 is directed towards C, so that the resulting motion is known as a sliding motion, with the relay theoretical switching at an infinite rate.

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. 2.6 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. If the hysteresis Δ is finite and h is large, then it is easily shown that a stable limit cycle, as shown in Fig. 2.7 for $\delta = 1$ and $\Delta = 0.5$, will occur. Trajectories both inside and outside the limit cycle have their motion directed towards it. Similarly, it is straightforward to draw phase-plane trajectories for a finite hysteresis Δ and smaller values of h.

Example 2

Fig. 2.8 shows the block diagram of an approximate model, since no viscous friction is assumed once again, of a position control system with nonlinear effects due to torque saturation and Coulomb friction.

The differential equation of motion in phase variable form is

$$\dot{x}_2 + f_s(-x_1) - \left(\frac{1}{2}\right) \operatorname{sgn}(x_2) = 0 \tag{2.6}$$

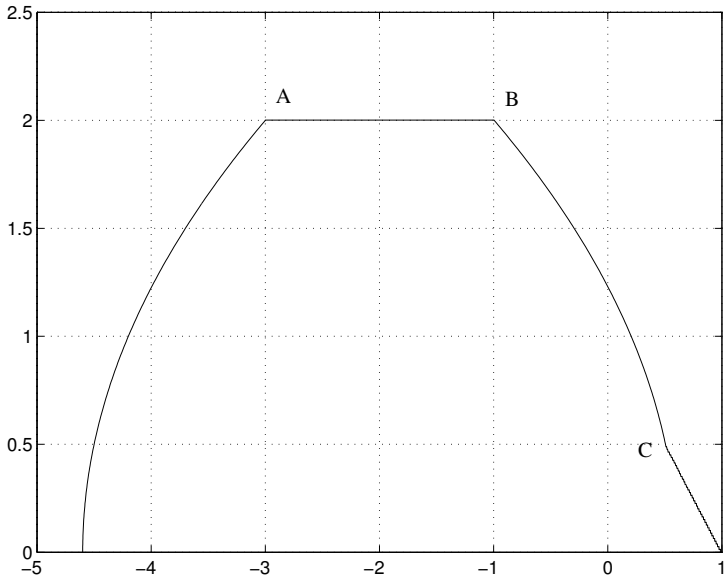

Figure 2.5: Response of system of Fig. 2.3 from an initial condition.

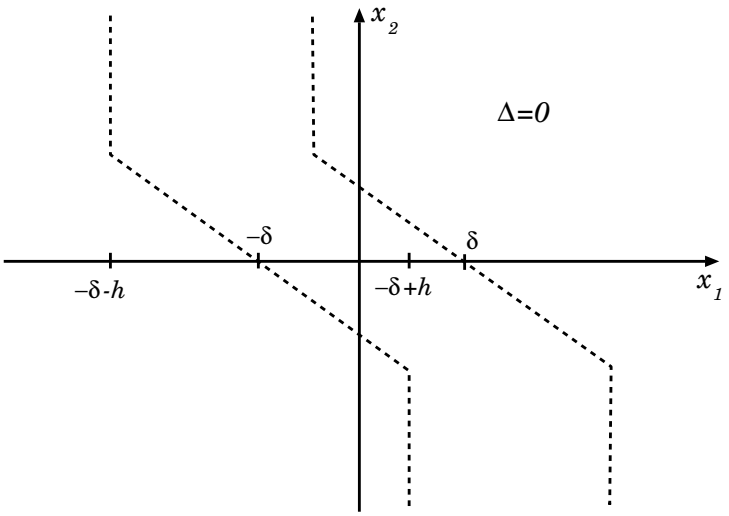

Figure 2.6: New switching boundaries with velocity saturation.

where f_s denotes the saturation nonlinearity and *sgn* the signum function, which is $+1$ for $x_2 > 0$ and -1 for $x_2 < 0$. There are six linear differential equations describing the motion in different regions of the phase plane. For x_2 positive, eqn 2.6 can be written

$$\ddot{x}_1 + f_s(x_1) + \frac{1}{2} = 0$$

so that for

(a) $x_2 + \nu e,\ x_1 < -2$, we have $\dot{x}_1 = x_2,\quad \dot{x}_2 = \frac{3}{2}$, a parabola in the phase plane.
(b) $x_2 + \nu e,\ |x_1| < 2$, we have $\dot{x}_1 = x_2,\quad \dot{x}_2 = \frac{3}{2}$, a circle in the phase lane.
(c) $x_2 + \nu e,\ x_1 > 2$, we have $\dot{x}_1 = x_2,\quad \dot{x}_2 = \frac{-5}{2}$, a parabola in the phase plane.

Similarly for x_2 negative,

(d) $x_2 - \nu e,\ x_1 < -2$, we have $\dot{x}_1 = x_2,\quad \dot{x}_2 = \frac{5}{2}$, a parabola in the phase plane.
(e) $x_2 - \nu e,\ |x_2| < 2$, we have $\dot{x}_1 = x_2,\quad \dot{x}_2 + x_1 - \frac{1}{2} = 0$, a circle in the phase lane.
(f) $x_2 - \nu e,\ x_1 > 2$, we have $\dot{x}_1 = x_2,\quad \dot{x}_2 = \frac{-3}{2}$, a parabola in the phase plane.

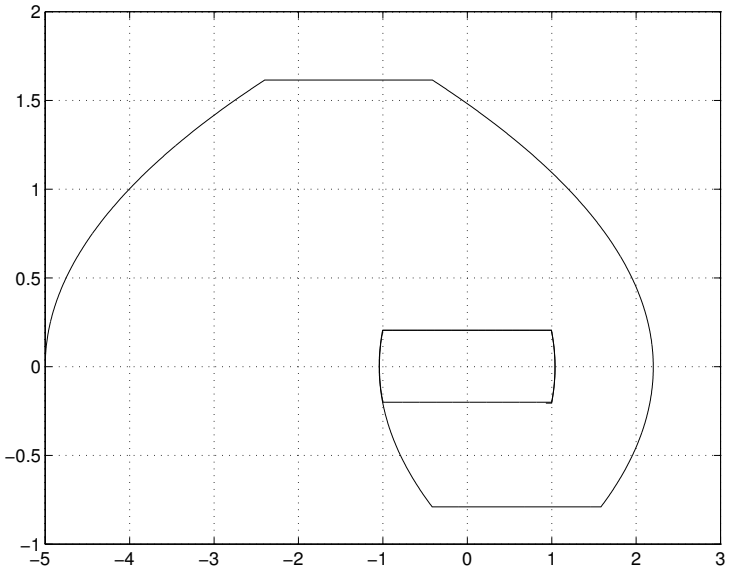

Figure 2.7: Limit cycle with velocity saturation.

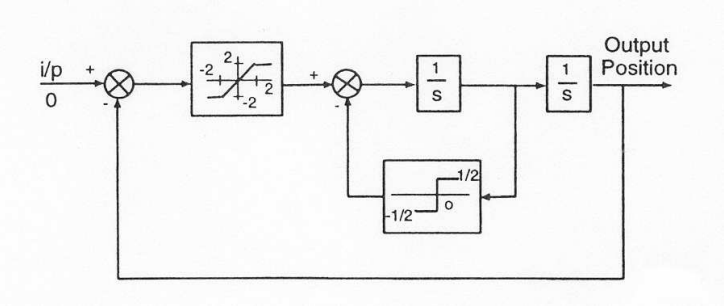

Figure 2.8: Block diagram of control system for Example 2.

Because all the phase plane trajectories are described by simple mathematical expressions, it is straightforward to calculate specific phase plane trajectories.

2.3 Concluding remarks

The above examples containing linear segmented nonlinear characteristics have been chosen to show (i) how the phase plane can be divided up into different regions in which the motion may be described by linear differential equations (ii) that when simple linear equations describe the motion analytical solutions may be found. The former concept is very useful to understand when doing simulations whereas the need for the latter is not required when simulations can be used. Several problems are given below which can be simulated in DYNAST to check your analytical solutions.

2.4 Problems

Problem 1.
An approximate block diagram for a position control system is shown in Fig. P1. It results in a closed loop transfer function which is second order, having a natural frequency of ω_0 and a damping ratio $\zeta = \zeta_V + \zeta_T$, where the former accounts for damping, such as viscous friction, and the latter the result of a velocity feedback stabilization signal. The effects of nonlinearity on the error and torque signals can be studied by placing the nonlinearity in e, and T, respectively. Simulate the linear system for $\theta_i = 0$, $\zeta = 1$ and any choice of ω_0. Show that, and also prove theoretically that, for initial conditions $\dot{\theta}_0 = v_0$ and $\theta_0 = -u_0$

1. there will be no overshoot if $v_0 < \omega_0 u_0$
2. the overshoot will be less than u_0 if $v_0 < 4.6\,\omega_0 u_0$.

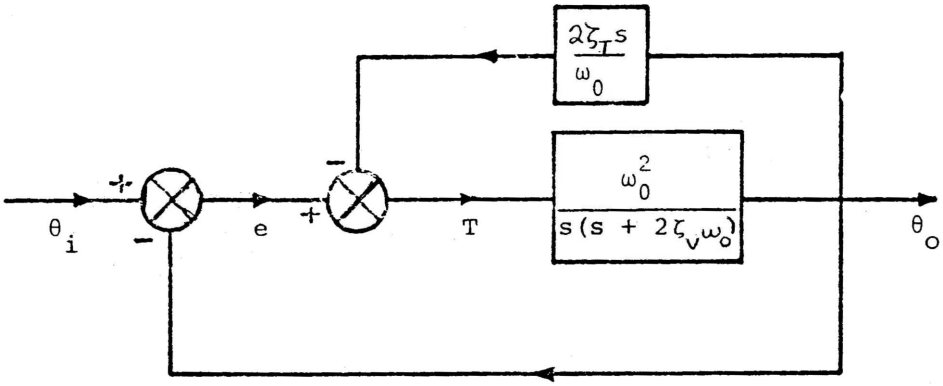

Figure P1.

Problem 2.
The system of Problem 1 has the ideal saturation characteristic of Figure 3.1, with linear slope m_e and saturation level h_e, in the error channel. Prove the following points if $\theta_i = 0$ and show that they are true by simulation

1. for initial conditions $\dot{\theta}_0 = v_0$ and $\theta_0 = -u_0$ with $u_0 > \delta_e$, where $h_e = m_e\delta_e$, the maximum velocity reached will be $h_e\omega_0/2\zeta_v$ provided v_0 is less than this value.
2. if $\zeta_T = 1$ and $m_e = 1$ show that the minimum value of ζ_v is 0.5 if no overshoot should occur for a large step input.

Problem 3.
Examine in your simulations for the linear system of Problem 1 the phase plane portrait for various initial conditions with $\zeta = 2$ and any choice of ω_0. You should note that the straight lines of slopes $(-2\pm\sqrt{3})\omega_0$, which are isoclines, are possible trajectories. Again for $\theta_i = 0$ and the initial conditions $\dot{\theta}_0 = v_0$ and $\theta_0 = -u_0$ show that there will be no overshoot provided $v_0 < (2+\sqrt{3})\omega_0$. What will be the ratio of the times required to travel to the origin for trajectories starting on the two isoclines with the same value of u_0.

Problem 4.
Simulate the system of Problem 1 with $\zeta_v = 0$, any choice of ω_0, and the ideal saturation characteristic of Figure 3.1 in the torque signal. Denote the parameters of the saturation by m_T, δ_T and h_T. Decrease δ_T to zero and do simulations for $\theta_i = 0$ and initial conditions $\dot{\theta}_0 = 0$ and $\theta_0 = -u_0$ for various values of u_0. Observe how changing the value of ζ_T changes the slope of the line where the trajectories switch from one parabolic form to another and for some choices of parameters that the final trajectory to the origin switches rapidly between the two parabolas (this is known as sliding). Show theoretically and by simulation that the maximum value of u_0 for no overshoot is $16\,\zeta_T^2 h_T$.

Problem 5.
For the previous example a designer decides that he needs to have critical damping for ζ. He estimates that ζ_v is 0.25. He therefore adjusts the velocity feedback gain to give ζ_T equal to 0.75. Simulate the system to determine the maximum value of u_0 for no overshoot. How much greater is this value than if ζ_v were zero.

Problem 6.
The system of Problem 1 has $\zeta_T = 0$ and Coulomb friction equal to the torque produced by a 4° error. Calculate the overshoots, undershoots and steady state error for a step input of 64° if $\zeta_v = 0.30$ and $\zeta_v = 0.15$. You may use the graph of Part III Figure 9.1.3. Check your results by simulation.

Problem 7.
For the system of Problem 4 a designer decides that for a step input the overshoot must be no greater

than 0.25 of the input magnitude. If $\zeta_T = 1$ calculate what will be the maximum allowable step input magnitude and check the result by simulation.

Problem 8.
A position control system has error saturation and a relay with dead zone as its actuator. It can be taken to have the block diagram of Figure 1.1 with $\omega_0 = 2$, $\zeta_v = 0$, $\zeta_T = 0.25$, error saturation with $m_e = 1$ and $\delta_e = 5°$, and the relay with $h = 4$, $\Delta = 0$ and $\delta = 1$. Sketch step responses on a phase plane for inputs of 10° and 20°, respectively. Check your results using simulation.

Problem 9.
The block diagram of a nonlinear control system is shown in Figure P9. Draw the switching boundaries on the phase plane and show the step responses on the phase plane for inputs of 6° and 10°, respectively. Sketch also the waveforms of $\dot{\theta}_0$ and θ_0 against time. State if the system is more or less damped for larger inputs and give your reasons. Check your results using simulation.

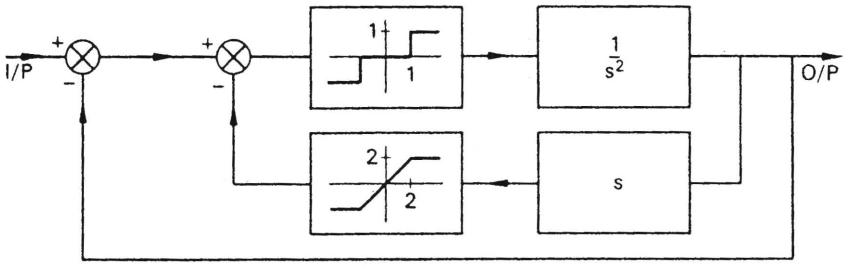

Figure P9.

Problem 10.
The block diagram of a relay control system with nonlinear velocity feedback is shown in Figure P10. Draw the switching boundaries on the phase plane and sketch some phase plane trajectories. In particular, determine the amplitude and frequency of any limit cycles. Check your results by simulation.

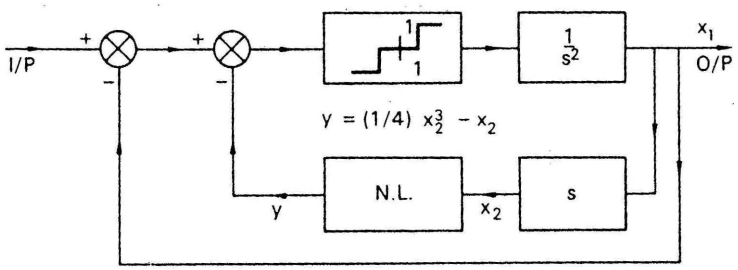

Figure P10.

Problem 11.
Sketch phase portraits for a relay position control system in the form of Figure P1 with $\zeta_v = 0$, $\zeta_T = 1$, $\omega_0 = 1$ and a relay in the torque channel with $h = 2$, $\Delta = 0.5$ and $\delta = 1.5$. Show that it has a limit cycle with a period of 24.5 seconds and check the result by simulation.

Module 3

The Describing Function

Module units

Module overview.

This module defines the describing function (DF) of a nonlinearity and shows how it may be used to study the behavior of a feedback loop with a single nonlinearity. DFs are derived for single and double valued nonlinearities and examples are given to show how DFs can be used for examining the stability of a nonlinear system. Since the method is approximate comments are made on its accuracy.

Module objectives. *When you have completed this module you should be able to:*

1. Define a describing function of a nonlinearity
2. Evaluate the describing function of a given nonlinearity
3. Use the describing function approach to understand the dynamic behavior of a feedback loop with a single nonlinearity
4. The approximations and limitations of the describing function method.

Module prerequisites. *First year mathematics, Fourier series,transfer functions,frequency response, stability*

3.1 Introduction

The describing function, which will be abbreviated DF, method was developed simultaneously in several countries during the wartime years of the early 1940s. Of major concern at that time was the design of control systems for tracking devices, such as gun and antenna systems. It was found that in some situations they would 'hunt', the name given to small amplitude limit cycles about the equilibrium point. It was realized that this instability was probably due to nonlinearities, such as friction or backlash in the gears of the control system, and an analysis method was required which could ensure that resulting designs would not exhibit this behavior. It was further observed that when limit cycles occurred the

motion of the system output was often approximately sinusoidal. This indicated to the engineers that a possible analytical approach would be to assume that the signal at the input to the nonlinear element in the loop was a sinusoid, the basis of the simple or sinusoidal DF approach. Since then there have been many developments in terms of both using the DF concept for other types of signals and the problems, or phenomena, which the DF can be used to study. Our considerations here will be primarily concerned with using the DF or S (sinusoidal) DF as it is often named.

Consider again the autonomous feedback system shown in Fig. 1.1 containing a single static nonlinearity $n(x)$ and linear dynamics given by the transfer function $G(s) = G_c(s)\,G_1(s)$. If a limit cycle exists in the autonomous system with the output $c(t)$ approximately sinusoidal, then the input $x(t)$ to the nonlinearity might also be expected to be sinusoidal. If this assumption is made the fundamental output of the nonlinearity can be calculated and conditions for the sinusoidal self-oscillation found, if the higher harmonics generated at the nonlinearity output are 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 the output fundamental may not be in phase with the sinusoidal input the DF may be complex.

3.2 Evaluation of the Describing Function

Assuming that in Fig. 1.1 $x(t) = a\cos\theta$, 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 $a_n = b_n = 0$ for n even, and in particular

$$a_1 = \frac{1}{\pi}\int_0^{2\pi} y(\theta)\cos\theta\, d\theta \tag{3.1}$$

and

$$b_1 = \frac{1}{\pi}\int_0^{2\pi} y(\theta)\sin\theta\, d\theta \tag{3.2}$$

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 $N_p(a) = a_1/a$ and $N_q(a) = -b_1/a$.

Alternatively, in polar co-ordinates, one can write $N(a) = M(a)e^{j\psi(a)}$, where $M(a) = (a_1^2 + b_1^2)^{1/2}/a$ and $\psi(a) = -\tan(b_1/a_1)$.

If $n(x)$ is single valued it is easily shown that $b_1 = 0$ and

$$a_1 = \frac{4}{\pi}\int_0^{\pi/2} y(\theta)\cos\theta\, d\theta \tag{3.3}$$

giving

$$N(a) = \frac{a_1}{a} = \frac{4}{a\pi}\int_0^{\pi/2} y(\theta)\cos\theta\, d\theta \tag{3.4}$$

Although eqns 3.1 and 3.2 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

$$a_1 = \frac{4}{a}\int_0^{a} x n_p(x) p(x)\, dx \tag{3.5}$$

$$b_1 = \frac{4}{a\pi}\int_0^{a} n_q(x)\, dx \tag{3.6}$$

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

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

and the nonlinear characteristics $n_p(x)$ and $n_q(x)$, called the in-phase and quadrature nonlinearities, are defined by

$$n_p(x) = [n_1(x) + n_2(x)]\,/2$$

and

$$n_q(x) = [n_2(x) - n_1(x)]\,/2$$

where $n_1(x)$ and $n_2(x)$ are the portions of a double valued characteristic traversed by the input for $\dot{x} > 0$ and $\dot{x} < 0$ respectively. For a single-valued characteristic, $n_1(x) = n_2(x)$, so that $n_p(x) = n(x)$ and $n_q(x) = 0$. Also integrating eqn. 3.5 by parts gives

$$a_1 = \frac{4}{\pi}\,n(0^+) + \frac{4}{a\pi}\int_0^a n'(x)(a^2 - x^2)^{1/2}\,dx \tag{3.7}$$

where $n'(x) = dn(x)/dx$ and $n(0^+) = \lim_{\varepsilon \to 0} n(\varepsilon)$. This is a useful expression for obtaining DFs for linear segmented characteristics. An additional advantage of using eqns 3.5 and 3.6 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 $0 < x < b$, then $N_\alpha(a) < N_\beta(a)$ for all input amplitudes a less than b.
3. For the sector bounded single-valued nonlinearity that is $k_1 x < n(x) < k_2(x)$ for all $0 < x < b$ then $k_1 < N(a) < k_2$ for all input amplitudes a 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:

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

3.3 DFs of some common nonlinearities

Tables of DFs for a variety of nonlinear characteristics can be found in many books. However, to illustrate the evaluation of the DF for some specific nonlinearities a few simple examples are considered below.

A cubic nonlinearity

For this nonlinearity $n(x) = x^3$ and using eqn. 3.3 one has

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

giving $N(a) = \frac{3a^2}{4}$.

Alternatively from eqn. 3.5

$$a_1 = \frac{4}{a}\int_0^a x^4 p(x)\,dx$$

The integral $\mu_n = \int_{-\infty}^{+\infty} x^n p(x)\,dx$ is known as the n^{th} 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

$$\mu_n = \begin{cases} 0 & \textit{for } n \textit{ odd} \\ a^n \frac{(n-1)}{n}\frac{n-3}{n-2}\ldots\frac{1}{2} & \textit{for } n \textit{ even} \end{cases}$$

Therefore $N(a) = \frac{4}{a^2}\frac{1}{2}\frac{3}{4}\frac{1}{2}a^4 = \frac{3a^2}{4}$ as before.

This nonlinearity is often referred to as a hard spring characteristic as its output increases more than linearly with the input. As expected, since the DF is a 'gain measure' this also increases with input amplitude.

An ideal saturation nonlinearity

Fig. 3.1 shows an ideal saturation characteristic together with the nonlinearity input, taken as $a\sin\theta$, and the corresponding output waveform $y(\theta)$.

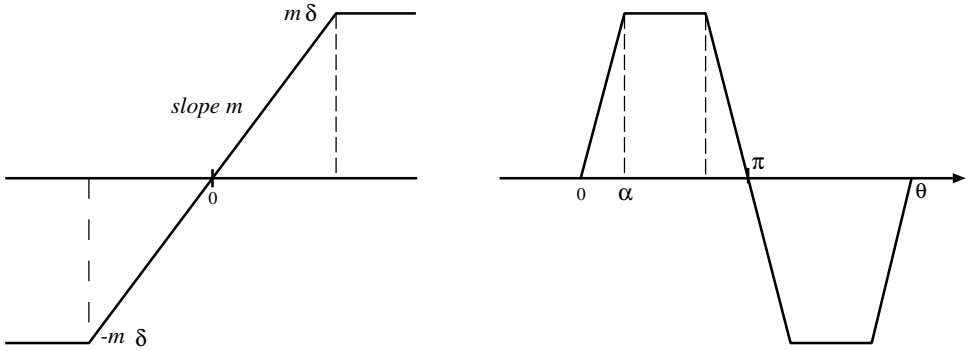

Figure 3.1: Ideal saturation and input and output waveforms.

Again, because of the symmetry of the nonlinearity the fundamental of the output can be evaluated from the integral over a quarter period so that

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

which for $a > \delta$ gives

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

and with $\alpha = sin^{-1}\delta/a$ evaluation of the integrals gives

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

which on substituting for δ gives the result

$$N(a) = \frac{m}{\pi}(2\alpha + \sin 2\alpha)$$

Since for $a < \delta$ the characteristic is linear giving $N(a) = m$, the DF for ideal saturation may be written as $mN_s(\delta/a)$ where

$$N_s(\delta/a) = \begin{cases} 1 & \textit{for} \quad a < \delta \\ (1/\pi)\left[2\alpha + \sin 2\alpha\right] & \textit{for} \quad a > \delta \end{cases}$$

Alternatively one can evaluate, $N(a)$, using eqn. 3.7, which yields

$$N(a) = \frac{a_1}{a} = \frac{4}{a^2\pi}\int_0^{\delta} m(a^2 - x^2)^{1/2}\,dx$$

Using the substitution $x = a\sin\theta$, this yields

$$N(a) = \frac{4m}{\pi}\int_0^{\alpha} \cos^2\theta\,d\theta = \frac{m}{\pi}(2\alpha + \sin 2\alpha), \text{ as before.}$$

This nonlinearity is often referred to as a soft spring characteristic and if the describing function, $N(a)$, is plotted against a it will be seen to decrease with a for $a > \delta$.

Relay with dead zone and hysteresis

The characteristic of a relay with dead zone and hysteresis is shown in Fig. 3.3 together with the input, assumed equal to $a\cos\theta$, and the corresponding output waveforms.

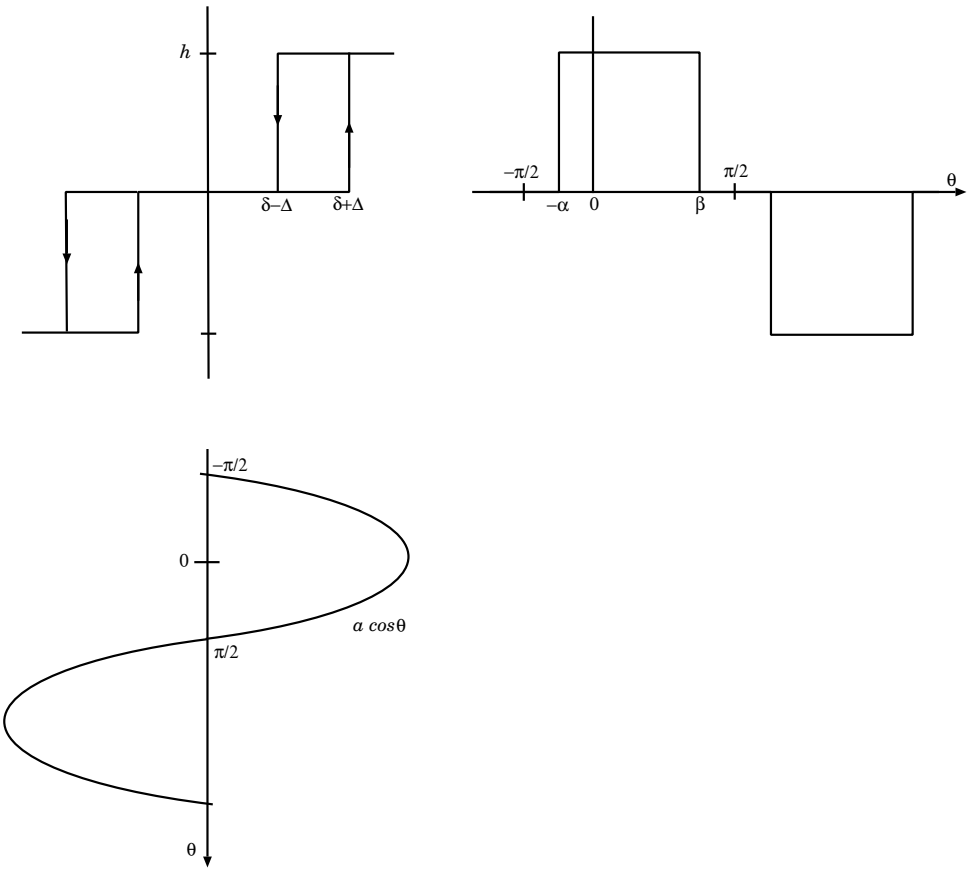

Figure 3.2: Relay with dead zone and hysteresis together with input and output waveforms.

Using eqns 3.1 and 3.2 over the interval $-\pi/2$ to $\pi/2$ and assuming that the input amplitude a is greater than $\delta + \Delta$, gives

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

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

$$\begin{aligned} b_1 &= \frac{2}{\pi}\int_{-\alpha}^{\beta} h\sin\theta\,d\theta = \frac{2h}{\pi}\left(\frac{\delta+\Delta}{a} - \frac{\delta-\Delta}{a}\right) = \frac{4h\Delta}{a\pi} \\ N(a) &= \frac{2h}{a^2\pi}\left\{\left[a^2 - (\delta+\Delta)^2\right]^{(1/2)} + \left[a^2 - (\delta-\Delta)^2\right]^{1/2}\right\} - \frac{j4h\Delta}{a^2\pi}. \end{aligned} \tag{3.8}$$

For the alternative approach one must first obtain the in-phase and quadrature nonlinearities which are shown in Fig. 3.3. Using eqns 3.5 and 3.6 one obtains

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

$$b_1 = \frac{4}{a\pi}\int_{\delta-\Delta}^{\delta+\Delta}\frac{h}{2}\,dx = \frac{4h\Delta}{a\pi} = \frac{\textit{Area of nonlinearity loop}}{a\pi}, \text{ as before}$$

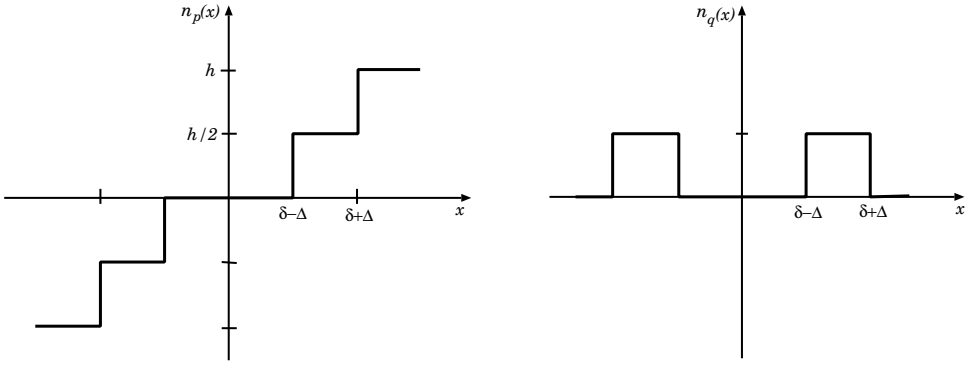

Figure 3.3: In-phase and quadrature nonlinearities.

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, $\Delta = \delta = 0$.

3.3.1 Further Comments

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. For example the DF of an ideal dead zone characteristic can easily be found from that for the ideal saturation of section 3.3, since an ideal dead zone with slope m outside the dead zone is simply a linear element of slope m – the ideal saturation, that is a gain of m in parallel with minus saturation. Several procedures 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 polynomial, a quantized characteristic, a linear segmented characteristic or a Fourier series.

3.4 Determination of the Stability of a Feedback Loop

Investigation of the stability of the nonlinear feedback loop of Fig. 1.1 using the DF method is based on the following fact. If $n(x)$ is an odd symmetrical characteristic lying within lines of slope k_1 and k_2, then the only form of instability the feedback loop can possess is a limit cycle provided that the linear system with $n(x)$ replaced by a gain K is stable for all gains between k_1 and k_2. To investigate the possibility of limit cycles in the autonomous closed loop system the nonlinearity $n(x)$ is replaced by its DF $N(a)$. In this case the open loop gain to a sinusoid is $N(a)\,G(j\omega)$ and a limit cycle will exist if

$$N(a)\,G(j\omega) = -1 \tag{3.9}$$

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. As $G(j\omega)$ is a complex function of ω and $N(a)$ may be a complex function of a, any solution to eqn. 3.9 will yield both the frequency ω and amplitude a of a possible limit cycle. Various approaches can be used to examine eqn. 3.9 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, when $n(x)$ is single-valued, 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 $s \approx j\omega$.

Fig. 3.4 illustrates the procedure on a Nyquist diagram, where sketches of possible $G(j\omega)$ and $C(a) = -1/N(a)$ loci are plotted and shown intersecting for $a = a_0$ and $\omega = \omega_0$

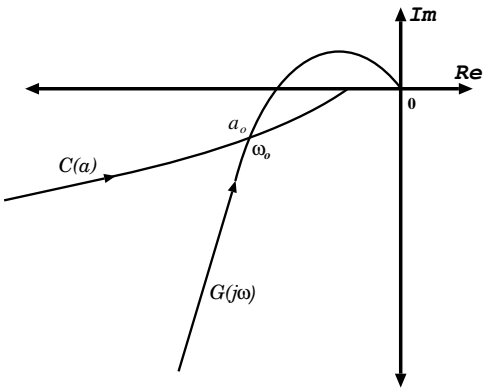

Figure 3.4: Limit cycle found from intersection of $G(j\omega)$ and C(a) on a Nyquist plot.

The DF method thus 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 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. If a solution is indicated then accurate values for a and ω can be found using a nonlinear algebraic equation solver.

A further point of interest when a solution exists to eqn. 3.9 is whether the predicted limit cycle is stable or unstable. This is obviously important if the control system is designed to have a limit cycle operation, as is the case for 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 continue to grow. The stability of a limit cycle, provided 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 stability (instability) for a point on $C(a)$ with $a > a_0$, then the limit cycle is stable (unstable). The situation may be 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.

3.5 DF Accuracy

A specification for many control systems is that when perturbed by either a step input or disturbance they should return within a finite time to a static equilibrium. For a nonlinear system this means checking that (a) the response will not go unbounded and (b) no limit cycle will occur. Condition (a) can be checked from linear methods as indicated previously and (b) by the DF method. The DF method, however, can only do this approximately since it assumes any limit cycle will be sinusoidal at the input to the nonlinearity, which will never be quite true in practice. This means that the DF method could indicate incorrectly either the existence or non-existence of a limit cycle. It is therefore important to have some idea of the validity and accuracy of any DF result.

If the DF method predicts a limit cycle then its validity can normally be checked by assuming this sinusoidal signal as the nonlinearity input and evaluating the percentage distortion, d, in the signal fed back to the nonlinearity, assuming the loop open at the nonlinearity input. The distortion, d, is defined by

$$d = \left[\frac{\text{M.S. value of signal} - \text{M.S. value of fundamental harmonic}}{\text{M.S. value of fundamental harmonic}} \right]^{1/2}$$

If the percentage distortion in this signal is less than 10% then the DF prediction should be valid. It has also been shown theoretically that the estimate provided by the DF for the limit cycle frequency, ω, will be more accurate than for the amplitude, a. Further, a is the estimate for the fundamental component in the limit cycle not the peak amplitude. When the DF method does not predict a limit cycle it will be correct if the $C(a)$ and $G(j\omega)$ loci are not near to intersecting. If this is not the case, however, one can assume a possible limit cycle at the nonlinearity input with amplitude a and frequency ω corresponding to the near intersection values of $C(a) = G(j\omega)$. Using the distortion criterion given above a very low value must exist for the prediction to be valid.

3.6 Concluding Remarks

The purpose of this section has been to provide a brief introduction to the sinusoidal DF and its use for the study of feedback loop stability. It has many more possibly applications details of which can be found in specialist texts. These include use in certain situations for loops with more than one nonlinearity, assistance in compensator design, and determination of the closed loop frequency response. Describing functions have been defined for other inputs and for multiple inputs which enable the approach to be used to study the interesting and unique behavior modes, mentioned in section 1.3, which can occur in nonlinear feedback systems. A few problems are given below where the DF calculated results can be compared with the results from DYNAST simulations

3.7 Problems

Problem 1.
Evaluate the amplitude at the output, θ_0, of the limit cycle in Problem 10 using the DF method. Compare the exact frequency of oscillation with that given by the DF method.

Problem 2.
Evaluate the amplitude at the output, θ_0, of the limit cycle in Problem 11 using the DF method. Compare the exact frequency of oscillation with that given by the DF method.

Problem 3.
Simulate the following two differential equations to find any limit cycles and compare your solutions with those given by the DF method.

1. $\ddot{x} - \dot{x} + 0.25\dot{x}|\dot{x}| + x = 0$
2. $\ddot{x} - 0.5\dot{x} + 2\dot{x}^3 + x = 0$

Module 4

Limit cycle and absolute stability

Module units

Module overview.

This module shows how limit cycles can be found in a feedback loop both approximately by using the DF method and exactly by the Tsypkin method when the nonlinearity is a relay. The problem of the absolute stability of a feedback loop containing a single nonlinearity is discussed and some criteria that provide sufficient but not necessary conditions are presented.

Module objectives. *When you have completed this module you should be able to:*

1. Find limit cycles approximately in a nonlinear feedback loop using the DF
2. Find limit cycles exactly in a feedback loop with a relay
3. Understand the principle of autotuning
4. Understand simple methods for finding the absolute stability of a nonlinear feedback loop

Module prerequisites. *First year mathematics, Fourier series, differential equations, transfer functions, frequency response, stability*

4.1 Determination of Limit Cycles

Two examples are given in the following sections to illustrate limit cycle determination using the DF method for a feedback loop containing a single nonlinear element. Both examples take the nonlinearity to be a relay. In section 4.4 the Tsypkin method is presented which allows the exact evaluation of limit cycles in relay systems. The first example takes a relay with dead zone as the nonlinear element and by using results from the Tsypkin method one is able to show that when the nonlinearity output contains a higher percentage of harmonics the DF solution becomes less accurate. The second example considers a more recent use of the DF method whereby the intentional introduction of a limit cycle in a feedback loop

is used to determine information about the process to be controlled in order to obtain suitable controller parameters. Although the method is discussed with relevance to its popular usage in PID controllers, it can also be used for other simple controllers, such as phase lead.

4.2 Feedback Loop Containing a Relay with Dead Zone

For this example the feedback loop of Fig. 1.1 is considered with $n(x)$ a relay with dead zone and $G(s) = 2/s(s+1)^2$. From eqn. 3.8 with $\Delta = 0$, the DF for this relay, given by $N(a) = 4h(a^2 - \delta^2)^{1/2}/a^2\pi$ for $a > \delta$, and is real because the nonlinearity is single valued. A graph of $N(a)$ against a is given in Fig. 4.1 and shows that $N(a)$ starts at zero, when $a = \delta$, increases to a maximum, with a value of $2h/\pi\delta$ at $a = \delta\sqrt{2}$, then decreases toward zero for larger inputs. The $C(a)$ locus, shown in Fig. 4.2, lies on the negative real axis starting at $-\infty$ and returning there after reaching a maximum value of $-\pi\delta/2h$. The frequency response of the given transfer function $G(j\omega)$ crosses the negative real axis, as shown in Fig. 4.2, at a frequency of $\tan^{-1}\omega = 45°$, that is $\omega = 1$ *rad/sec* , and therefore cuts the $C(a)$ locus twice.

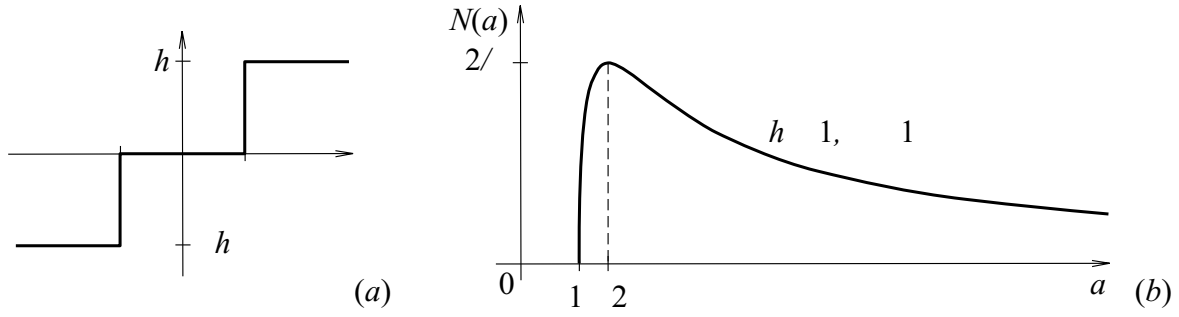

Figure 4.1: Ideal relay with dead zone and its DF.

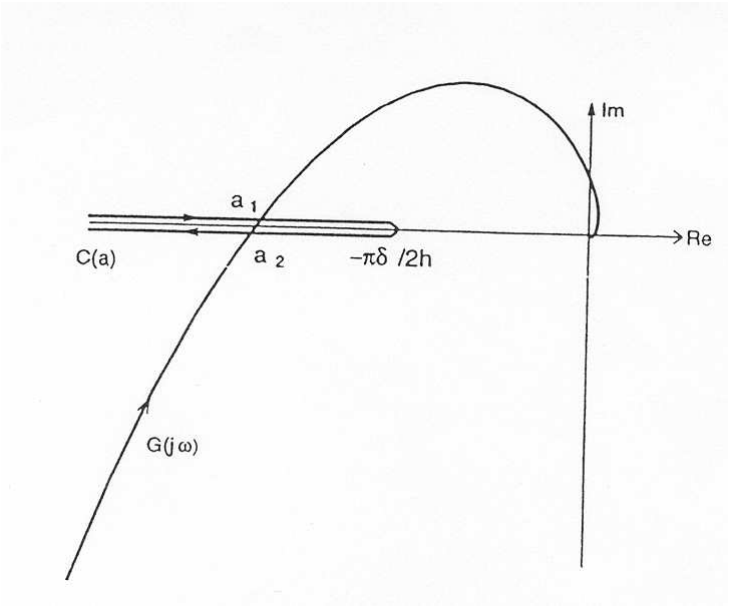

Figure 4.2: Solutions for the two limit cycles, a_1 unstable; a_2 stable.

The two possible limit cycle amplitudes at this frequency can be found by solving

$$\frac{a^2\pi}{4h(a^2 - \delta^2)^{1/2}} = 1$$

which gives $a = 1.04$ and 3.86 for $\delta = 1$ and $h = \pi$. Using the stability criterion given in section 3.4 it is seen that the smallest amplitude limit cycle is unstable and the larger one is stable. If a condition similar to the lower amplitude limit cycle is excited in the system, an oscillation will build up and stabilize at

the higher amplitude limit cycle. The exact frequencies of the limit cycles for the smaller and larger amplitudes are 0.709 and 0.989 respectively. Although the transfer function is a good low pass filter, the frequency of the smallest amplitude (unstable) limit cycle is not predicted accurately because the output from the relay, a waveform with narrow pulses, is more highly distorted.

If the transfer function of $G(s)$ is $K/s(s+1)^2$, then no limit cycle will exist in the feedback loop, and it will be stable if

$$\left.\frac{K}{\omega(1+\omega^2)}\right|_{\omega=1} < \frac{\pi d}{2h}$$

that is, $K < \pi\delta/h$. If $\delta = 1$ and $h = \pi$, $K < 1$ which may be compared with the exact result for stability of $K < 0.96$.

4.3 Autotuning in Process Control

In 1943 Ziegler and Nichols (Z-N), as discussed in Part III section 8.4, suggested two approaches for tuning the parameters of a PID controller. The first method described, the loop cycling method, is based on testing the plant in a closed loop with the PID controller in the proportional mode. The proportional gain is increased until the loop starts to oscillate and then the value of gain and the oscillation frequency are measured. Formulae were given, see Part III Table 8.4.1 for setting the controller parameters based on the gain, named the critical gain, K_c, and the frequency called the critical frequency, ω_c. There are basically two aspects to this method (a) finding K_c and ω_c , and then (b) deciding how to use this information about the plant in setting the controller parameters. In recent years many further formulae have been given for setting the controller parameters from knowledge of ω_c and K_c than the one given by Z-N.

Assuming that the plant has a linear transfer function $G_1(s)$, then K_c is its gain margin and ω_c the frequency at which its phase shift is 180°. Performing the Z-N test in practice to find the parameters can prove difficult. If, for example, the plant has a linear transfer function and the gain is adjusted too quickly, a large amplitude oscillation may start to build up. Also the oscillation amplitude will never stabilize due to the presence of noise unless some nonlinear operation takes place in the loop. In 1984 Astrom and Hagglund therefore suggested obtaining the required information by replacing the proportional control by a relay element to control the amplitude of the oscillation, strictly limit cycle.

Consider therefore the feedback loop of Fig. 1.1 with $n(x)$ an ideal relay, $G_c(s) = 1$, and the plant with a transfer function $G_1(s) = 10/(s+1)^3$. The $C(a)$ locus, $-1/N(a) = -a\pi/4h$, and the Nyquist locus $G(j\omega)$ intersect as shown in Fig. 4.3. The values of a and ω at the intersection can be calculated from

$$\begin{aligned} -\frac{a\pi}{4h} &= \frac{10}{(1+j\omega)^3}, \quad \text{which can be written} \\ \text{Arg}\ \frac{10}{(1+j\omega)^3} &= 180^\circ \quad \text{and} \quad \frac{a\pi}{4h} = \frac{10}{(1+\omega^2)^{3/2}}. \end{aligned} \tag{4.1}$$

The solutions for ω_c from eqns 4.1 is $\tan^{-1}\omega_c = 60^\circ$, giving $\omega_c = \sqrt{3}$, and for a one has $a = 5h/\pi$, giving the gain through the relay $N(a)$, which is the critical gain $K_c = 0.8$. Because the DF solution is approximate, the actual measured frequency of oscillation will differ from this value by an amount which will be smaller the closer the oscillation is to a sinusoid. The exact frequency of oscillation, which will of course be the measured value, is 1.708 *rads/sec* which differs from ω_c by a relatively small amount. For a square wave input to the plant at this frequency, the plant output signal will be distorted by a small percentage.

In the practical situation where a is measured, K_c is then calculated using the relay DF expression, that is $K_c = 4h/a\pi$. Although a should be the fundamental of the waveform the single measurement of peak amplitude, A, is normally made so that a is taken equal to A. In this example the distortion is very small and the accuracy of the estimates obtained for K_c and ω_c will be more affected by measurement errors on noisy signals than the error produced in using their values in the DF approximation equations.

If the relay has an hysteresis of Δ, then putting $\delta = 0$ in eqn. 3.8 gives

$$N(a) = \frac{4h(a^2-\Delta^2)^{1/2}}{a^2\pi} - j\frac{4h\Delta}{a^2\pi}$$

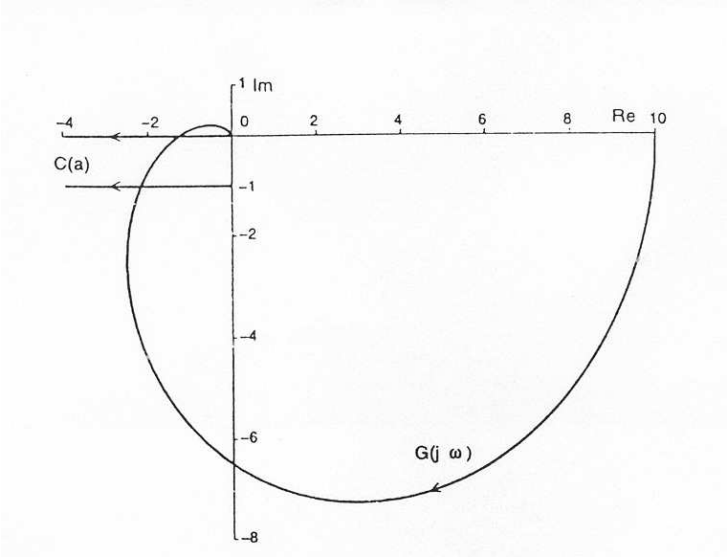

Figure 4.3: Nyquist plot for $10/(s+1)^3$ and $C(a)$ loci for $\Delta = 0$ and $4h/\pi$.

from which

$$C(a) \quad = \quad \frac{-1}{N(a)} = -\frac{\pi}{4h}\left[(a^2-\Delta^2)^{1/2} + j\Delta\right]. \tag{4.2}$$

Thus on the Nyquist plot, $C(a)$ is a line parallel to the real axis at a distance $\pi\Delta/4h$ below it, as shown in Fig. 4.3 for $\Delta = 1$ and $h = \pi/4$ giving $C(a) = -(a^2-1)^{1/2} - j$. If the same transfer function is used for the plant, then the limit cycle solution is given by

$$-(a^2-1)^{1/2} - j = \frac{10}{(1+j\omega)^3}$$

where $\omega = 1.266$, which compares with an exact solution value of 1.254, and $a = 1.91$. It is easily shown from the stability criterion that these limit cycles are stable.

4.4 The exact evaluation of limit cycles in relay systems

In this section an exact method for the evaluation of limit cycles and their stability in a feedback loop with a relay is discussed. The method makes use of the fact that the output from a relay is not continuously affected by its input. 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, which contrasts with the DF method where an approximate nonlinearity input waveform is assumed. The output waveform, $y(t)$, for a relay with dead zone and hysteresis, as shown in Fig. 3.2, may be assumed to take the form shown in Fig. 4.4, where T and Δt are unknown and the initial relay switching from 0 to h takes place at time t_1.

Then to find a possible limit cycle in the autonomous system of Fig. 1.1, where $n(x)$ is the relay and $G(s) = G_c(s)G_1(s)$, the steady state response of $G(s)$ to this waveform has to be determined. Several slightly different approaches are possible but here, that used by Tsypkin, primarily because for a relay with no dead zone it allows a simple comparison with the DF method, is presented. If $y(t)$ is expanded in a Fourier series then one obtains

$$y(t) = \frac{2h}{\pi}\sum_{n=1(2)}^{\infty}\frac{1}{n}\left\{\sin(n\omega\Delta t)\cos[n\omega(t-t_1)] + [1-\cos(n\omega\Delta t)\sin[n\omega(t-t_1)]]\right\}$$

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

$$c(t) = \frac{2h}{\pi}\sum_{n=1(2)}^{\infty}\frac{g_n}{n}\left\{\sin(n\omega\Delta t)\cos\left[n\omega(t-t_1)+\phi_n\right] + [1-\cos(n\omega\Delta t)\sin\left[n\omega(t-t_1)+\phi_n\right]]\right\}$$

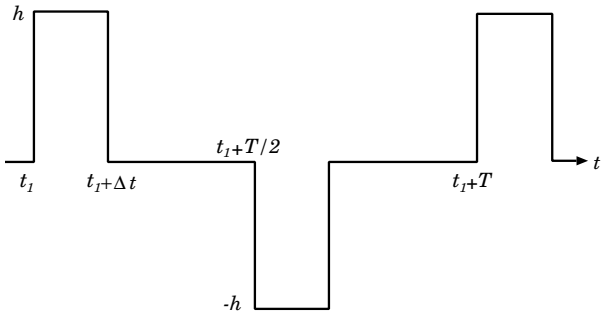

Figure 4.4: Relay output waveform.

where

$$g_n = |G(j\omega n)| \text{ and } \phi_n = \angle G(j\omega n)$$

Using the A loci, defined by

$$\begin{aligned} A_g(\theta,\omega) &= \mathrm{Re}\, A_g(\theta,\omega) + j\mathrm{Im}\, A_g(\theta,\omega) \\ \mathrm{Re}\, A_g(\theta,\omega) &= \sum_{n=1(2)}^{\infty} V_g(n\omega)\sin(n\theta) + U_g(n\omega)\cos(n\theta) \qquad (4.3) \\ \mathrm{Im}\, A_g(\theta,\omega) &= \sum_{n=1(2)}^{\infty} \frac{1}{n}\{V_g(n\omega)\cos(n\theta) - V_g(n\omega)\sin(n\theta)\} \qquad (4.4) \end{aligned}$$

where $U_G(n\omega) = g_n \cos\phi_n$ and $V_G(n\omega) = g_m \sin\phi_n$, then the expression for $c(t)$ 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, $\dot{c}(t)$ 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 for the autonomous system $x(t) = -c(t), \dot{x}(t_1)$, should be positive and $\dot{x}(t_1 + \Delta t)$ negative, it is necessary that

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

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 $T = 2\pi/\omega$ 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

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

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 eqns 4.3 and 4.4. 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, although this will not be pursued here.

4.5 Two Examples

As applications of this method two simple examples are considered.

Relay with no dead zone

Consider a relay with no dead zone, that is $\delta = 0$, so that one has the single relationship

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

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 to note that, since the line with $RP < 0$ and $IP = -\pi\Delta/4h$ corresponds to $C(a)$, the negative reciprocal of the DF, the exact solution and approximate DF solution can be compared graphically. This is done in Fig. 4.5, 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 $1.365\,rad\,s^{-1}$, and the approximate solution using the DF method is $1.352\,rad\,s^{-1}$.

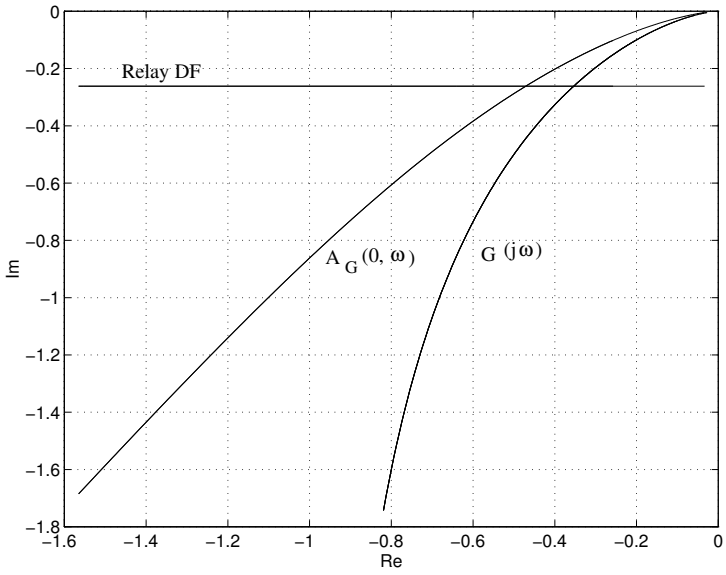

Figure 4.5: Graphical solution for limit cycle from $G(j\omega)$ and $A_G(0,\omega)$ loci.

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.

Relay with dead zone

Consider as a second example a feedback system having a relay with output ± 1, $\delta = 1$ and $\Delta = 0$, and 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. 4.6 and 4.7 were found. The larger amplitude limit cycle of Fig. 4.6 is shown 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. 4.7 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 $1.000\,rads/s$.

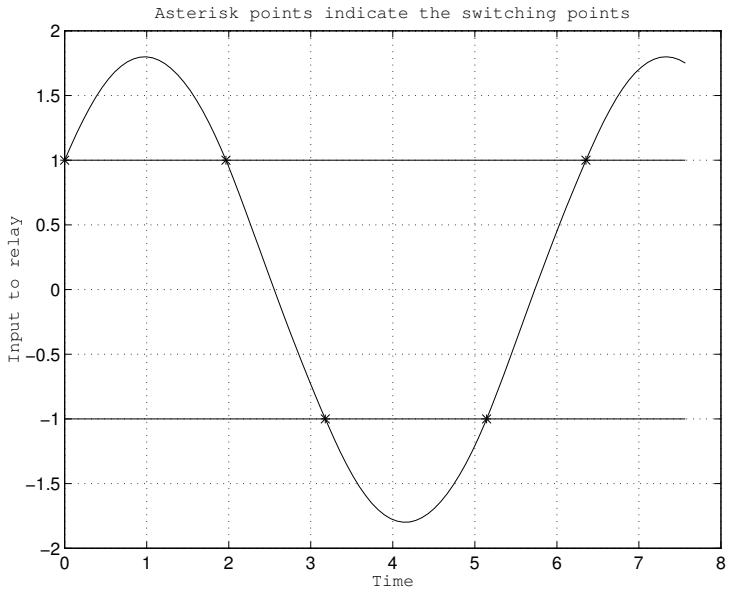

Figure 4.6: Graph of larger limit cycle solution.

4.6 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. 1.1, as described in Part III module 5, 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. 1.1 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 and further 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 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 $k_1\, x$ and $k_2\, x$ if $k_1 x^2 < xn(x) < k_2 x^2$ for all x. Similarly it has slope bounds (k_1', k_2') if $k_1' x^2 < xn'(x) < k_2' x^2$, where $n'(x) = dn(x)/dx$. The Popov criterion states that a sufficient condition for the autonomous system of Fig 1.1 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

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

where the nonlinearity $n(x)$ lies in the sector $(0, k)$. The theorem has the simple graphical interpretation shown in Fig. 4.8 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

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

The circle criterion can be obtained from the Popov criterion but its validity, using different analytical approaches, has been extended to cover the situation of a bounded input, r, to the system of Fig 1.1. 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

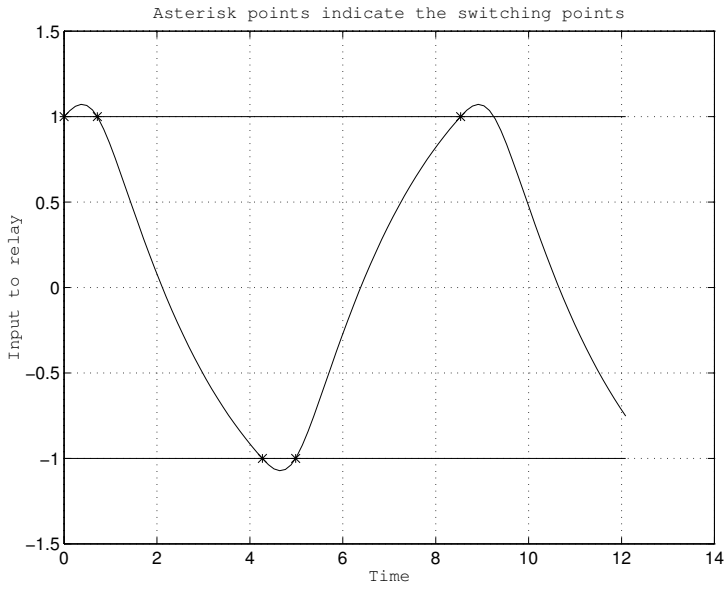

Figure 4.7: Graph of smaller limit cycle solution.

stability of the system of Fig. 1.1 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 for *(i)* $k_1k_2 < 0, G(j\omega)$ should be entirely within C, *(ii)* if $k_1k_2 > 0,\ G(j\omega)$ should lie entirely outside and not encircle C and *(iii)* 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 *(ii)* is shown in Fig. 4.9.

Two simple transformations are also useful for investigating the absolute stability of the autonomous system of Fig. 1.1, assuming $G_c(s) = 1$ and $G(s) = G_1(s)$. 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 the linear transfer function to $\rho + G(s)$.

Prior to the derivation of these frequency domain results Aizermann had put forward a conjecture that the autonomous system of Fig. 1.1 would be stable for a nonlinearity sector bounded by (k_1, k_2) if for k_1 and $k_2 > 0$ 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.4.9 and also the bounds for the DF of any single-valued nonlinearity within the sector. Several counter examples 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. For a monotonic nonlinearity with slope bounds (k_1', k_2') and $k_1'\, k_2' > 0$ an off axis circle criterion exists. This states that the autonomous system of Fig. 1.1 with a nonlinearity satisfying the aforementioned conditions will be absolutely stable if the Nyquist locus of a stable transfer function does not encircle any circle centered off the real axis and which intercepts it at $(-1/k_2', -1/k_2')$.

There are two viewpoints on the above absolute stability criteria. First, those of the engineer, who may argue that the results the criteria produce are too conservative if he wishes to apply them to a specific system with a well-defined nonlinearity for which much more is known about it than its sector bounds. Second, those of the theoretician, who may argue that the results are very robust in the sense that they guarantee stability for a system with a poorly defined nonlinearity, that is any nonlinearity satisfying certain sector properties.

4.7 Conclusions on Part IV

The purpose of Part IV has been to provide a brief introduction to nonlinearity and some methods for the analysis of a nonlinear feedback system with a single, or possibly more, nonlinearities. The methods that have been covered are those which should be easily understandable based on Part III of the course. To a large extent they may be regarded as simple extensions of the basic linear control ideas presented in Part

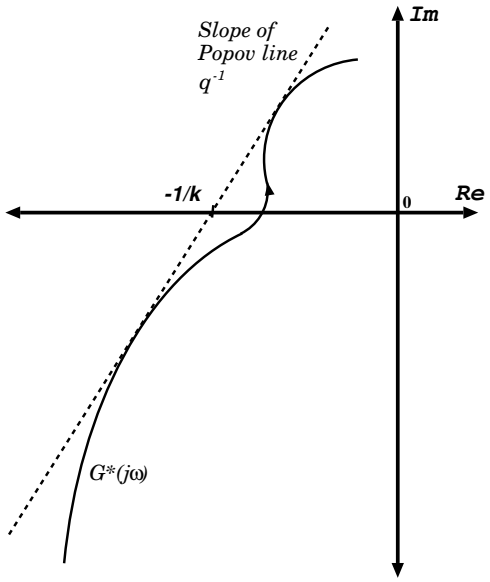

Figure 4.8: Graphical interpretation of Popov criterion.

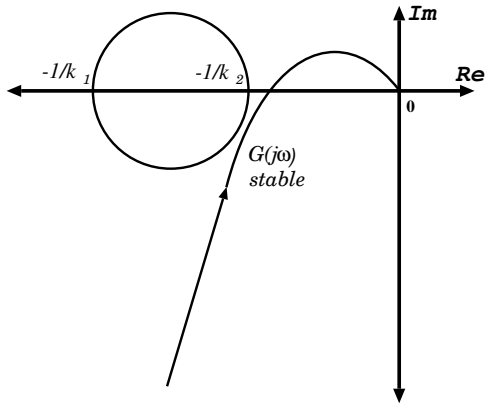

Figure 4.9: Illustration of the circle criterion.

III to enable some understanding of the effect nonlinearity might have on a simple feedback loop. Further the use of simulation studies provides an excellent approach to understanding the principles involved and their range of applicability. To assist the reader in this way several problems are given in for which the theoretical results can be checked by simulations in DYNAST. To learn more of the methods presented in Part IV and many others the reader is referred to the books given in the bibliography.

4.8 Problems

Problem 1.
For the autotuning procedure described in section 3.3 consider an ideal relay with output ± 1 and a plant with transfer function $\frac{4}{s}e^{-s}$. Simulate the system and find the amplitude and frequency of the limit cycle. Now calculate their values using the DF method. Why does the DF method despite a triangular rather than sinusoidal oscillation give the correct frequency. Remembering that the DF method predicts the fundamental of the oscillation amplitude calculate the corresponding amplitude of a triangular waveform. Does this agree with the measured amplitude?

Problem 2.
Consider again the autotuning experiment of Problem 1 but change the plant transfer function to $\frac{4e^{-s\tau}}{1+sT}$. In order to study autotuning for this process you can normalize the frequency by taking a new 's' equal to

sT and get the normalized transfer function $\frac{4e^{-s\rho}}{1+s}$ where $\rho = \tau/T$. Simulate the system for ρ in the range 0.1 to 10. Note how the waveform of the limit cycle changes. Plot a graph of the DF method solution for the frequency of oscillation as a function of ρ and mark on it some measured values for comparison. Explain the results. How could you get better estimates from the DF method for the oscillation amplitude for small and large ρ.

Problem 3.
An electromechanical system has a transfer function of $\frac{1}{s(1+s)(1+sT)}$. Set up an autotuning simulation with an ideal relay and use the measured oscillation amplitude to calculate the gain margin for values of T in the range 0.1 to 10. Plot a graph of the exact calculated values for comparison.

Problem 4.
A simple feedback control system has unity negative feedback and a forward path consisting of a symmetrical odd nonlinear element and a plant with transfer function $\frac{1}{s(1+s)^2}$. The nonlinear element is linear segmented having a slope of 0.5 from the origin to $x = 0.6$, a slope of 4 from $x = 0.6$ to $x = 0.7$, and a slope of 1 thereafter. A constant disturbance exists at the plant input. Simulate the system for step inputs of 4 units as the value of the constant disturbance is increased from 0 to unity. Explain why a limit cycle exists over a range of the constant disturbance values.

Problem 5.
Simulate an autotuning experiment with an ideal relay for the following three transfer functions:

(i) $\frac{4}{(s+1)^3}$

(ii) $\frac{1}{s(s+1)^2}$

(iii) $\frac{2}{s(s+1)(s+4)}$

Determine the critical frequency and critical gain from the measured results. Compare these with the calculated values and explain why they are in good agreement.

Problem 6.
Repeat the simulations of Problem 5 but this time using a relay with hysteresis. Calculate using the DF method the limit cycle frequency for all three transfer functions for the relay having Δ/h ratio of (i) 0.2 and (ii) 0.4. Compare your results with the measured values from the simulations.

Problem 7.
In doing autotuning testing if there is time to experiment then one can always get additional or better information by adding additional transfer functions to the loop. Here the Problem is to get better estimates for the critical frequency and critical gain of the normalized transfer function of Problem 2 with $\rho = 1$. Do this by adding a filter with transfer function $\frac{s\omega_0}{s^2 + 0.2s\omega_0 + \omega_0^2}$ before the relay. If you guess the correct value of the critical frequency for ω_0 then there will be no phase shift through the filter and the limit cycle will be almost sinusoidal. For typical process plant transfer functions it can be shown that the critical frequency will be slightly higher than the measured limit cycle frequency. Use this information to choose ω_0 after performing the autotuning test and then do the autotuning simulations with the filter. See how many iterations it takes you to get ω_0 and the critical gain to within 1% of their exact values.

Problem 8.
Describe how you would perform an autotuning test to obtain the frequency and corresponding gain for where the transfer function $\frac{1}{(s+1)(s+4)}$ has a phase shift of 90°. Do the simulation and compare your answers with the exact values.

4.9 Bibliography

[1] Astrom K. J. and Hagglund T. (1984). Automatic tuning of simple regulators with specifications on phase and amplitude margins. *Automatica*, Vol. 20, No. 5, pp 645–651

[2] Atherton, D. P. (1975). *Nonlinear Control Engineering: Describing Function Analysis and Design.* London: Van Nostrand Reinhold, 1975

[3] Atherton D. P. (1981). *Stability of Nonlinear Systems.* Research Studies Press, John Wiley

[4] Balasubramanian, R. (1981). Stability of limit cycles in feedback systems containing a relay. *IEE Proc. D*, No. 1, pp 24-29

[5] Blaquiere, A. (1966). *Nonlinear Systems Analysis.* New York: Academic Press

[6] Choudhury, S. K. and Atherton, D. P. (1974). Limit cycles in high-order nonlinear systems. *Proc. Inst. Electr. Eng.* 121, pp 717–724

[7] Cook P. A. (1994). *Nonlinear Dynamical Systems.* London: Prentice-Hall International, 2nd ed

[8] Cosgriff, R. L. (1958). *Nonlinear Control Systems.* New York: McGraw Hill

[9] Cunningham, W. J. (1958). *Introduction to Nonlinear Analysis.* New York: McGraw Hill

[10] Friedland, B. (1996). *Advanced Control System Design,* Prentice Hall.

[11] Gelb, A. and Vander Velde, W. E. (1968). *Multiple-Input Describing Functions and Nonlinear System Design.* New York: McGraw Hill

[12] Gibson, J. E. (1963). *Nonlinear Automatic Control.* New York: McGraw Hill

[13] Graham, D. and McRuer, D. (1961). *Analysis of Nonlinear Control Systems.* New York: Wiley

[14] Hayashi, C. (1964). *Nonlinear Oscillations in Physical Systems.* New York: McGraw Hill Kalman, R. E. (1954). Phase plane analysis of automatic control systems with nonlinear gain elements. *Trans. AIEE,* 73(II), pp 383–390

[15] Kalman, R. E. (1955). Analysis and design principles of second and higher order saturating servomechanisms. *Trans. AIEE*, 74(II), pp 294–308

[16] Khalil, H.K. (1992) *Nonlinear Systems.* Macmillan, New York.

[17] McNamara O. P. and Atherton D. P. (1987). Limit Cycle Prediction in Free Structured Nonlinear Systems, *IFAC Congress, Munich*, Vol. 8, pp 23–28

[18] Minorsky, N. (1962). *Nonlinear Oscillations.* New York: Van Nostrand

[19] Mohler, R. R. (1991) Nonlinear Systems; Vol 1 Dynamics and Control, Prentice-Hall.

[20] Slotine J. J. E. and Li W. (1991). *Applied Nonlinear Control.* Prentice-Hall International

[21] Taylor, J. H. (1980). Applications of a general limit cycle analysis method for multivariable systems, in Ramnath, R. V., Hedrick, J. K. and Paynter, H. M (eds) *Nonlinear System Analysis and Synthesis.* New York: ASME, Vol. 2

[22] Thaler, G. J. and Pastel, M. P. (1962). *Analysis and Design of Nonlinear Feedback Control Systems.* New York: McGraw Hill

[23] Van der Pol, B. (1934). Nonlinear theory of electric oscillations. *Proc. IRE*, 22, pp 1051–1086

[24] Vidyasagar, M (1978) *Nonlinear Systems Analysis.* Prentice-Hall International

[25] West, J. C. (1960). *Analytical Techniques for Nonlinear Control Systems.* London: EUP

[26] West, J. C., Douce, J. L. and Naylor, R. (1954). The effects of some nonlinear elements on the transient performance of a simple r.p.c. system possessing torque limitation. *Proc. IEE*, 101, pp 156–165

[27] Ziegler J. G. and Nichols N. B. (1942). Optimum settings for automatic controllers. *Trans. ASME*, Vol. 64, pp 759–768