8 NONLINEAR SYSTEMS

We introduce here a few concepts and analysis techniques for nonlinear systems. The analysis and control of linear systems is a necessary step in understanding nonlinear dynamics. Although, as we have seen, almost every nonlinear system can be locally approximated by a linearized system, this corellation should not be pushed too far. For nonlinear systems the principle of superposition of solutions does not hold. There are no separate natural and forced motions. Twice the input does not mean twice the output. For nonlinear systems there may be a significant dependence of the response on the magnitude and type of the excitation. For example, a nonlinear system may have completely different behavior under step inputs of different magnitude, or sinusoidal inputs of different frequencies. The response may also depend drastically on the initial conditions. In fact for some systems it may happen that the long term behavior of the solutions may be effectively random, even though both the system and the input are purely deterministic, as a result of extreme sensitivity to initial conditions. Since one can never be exactly certain about the initial state, the final state of such a system may very well be unpredictable. Such essentially unpredictable deterministic systems are known as chaotic systems.

8.1 Introduction

As a first example of what may happen when nonlinearities are present in a physical system, consider the so called Duffing's equation. This is nothing but a spring–mass–damper system with nonlinear spring force characteristics,

$$m\ddot{x} + b\dot{x} + kx + \alpha x^3 = 0 \ .$$

The spring force is $kx + \alpha x^3$ instead of kx that would be if the spring were linear. We call the case of $\alpha > 0$ a hardening spring, and $\alpha < 0$ a softening spring. A typical example would be the familiar $GZ(\phi)$ curve: it has the characteristics of a hardening spring for small ϕ for a surface ship, and a softening spring for a submarine. The plot of spring force vs. spring displacement would typically appear as shown in Figure 39.

We know that the natural frequency of oscilation of the linear spring system is $\omega_n = \sqrt{k/m}$, in other words it depends only on k and not on the amplitude of oscillation. For a hardening spring, it can be seen that the equivalent linearized spring constant is $k + 3\alpha x^2$, which means that it increases with the displacement x. Therefore, we expect the natural frequency of the hardening spring system to increase with the amplitude of oscillation, as well. The opposite is true for the softening spring case, $\alpha < 0$, see Figure 40.

Now consider Duffing's equation with forcing,

$$m\ddot{x} + b\dot{x} + kx + \alpha x^3 = P \cos \omega t \ .$$

We know that the frequency response curve has the familiar shape of Figure 41. It starts from 1, it may reach a maximum at about ω_n depending on the amount of damping, and

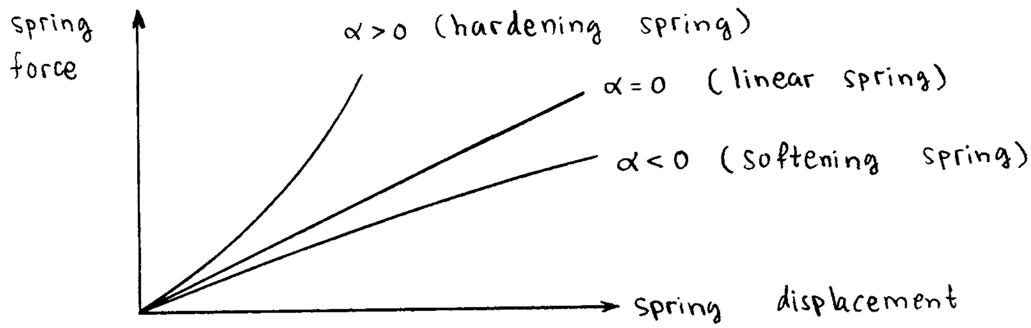

Figure 39: Linear and nonlinear spring force/displacement characteristics

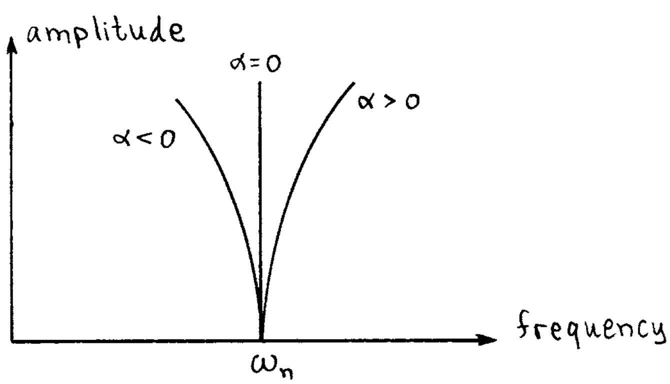

Figure 40: Natural frequency of linear and nonlinear springs

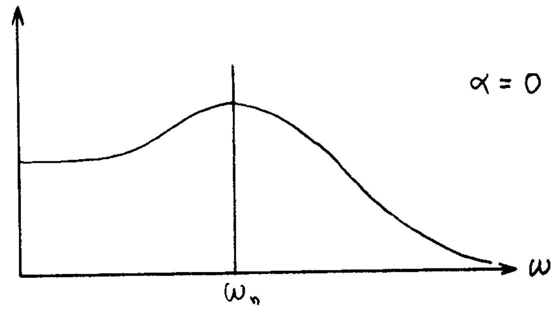

Figure 41: Frequency response curve for a linear spring

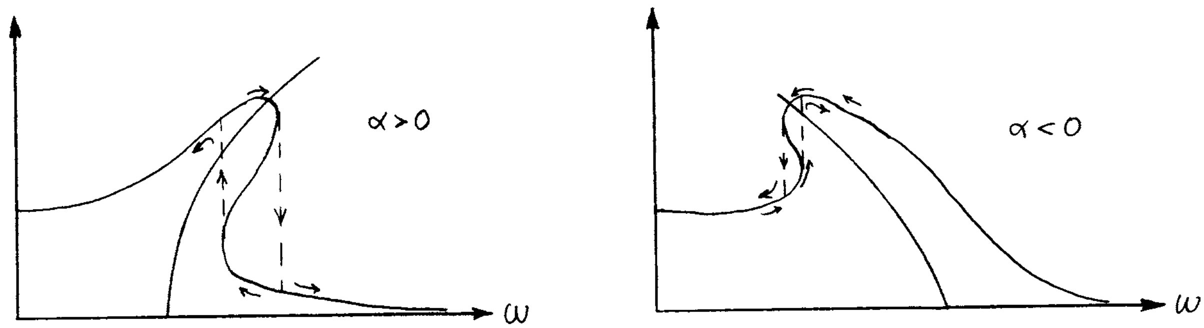

Figure 42: Frequency response curves for nonlinear springs

then it approaches zero. We can observe that the frequency response curve “wraps around” the amplitude vs. frequency curve we had before. Therefore, we can guess that the frequency response curves for the hardening and softening nonlinear springs will take one of the two forms shown in Figure 42.

We can see that depending on the frequency of excitation and upon increasing or decreasing this frequency, the system may experience oscillations with different amplitude, or sudden changes in the amplitude of the response. These phenomena are characteristic of nonlinear restoring forces and moments, and are called jump phenomena or hysteresis.

A different type of phenomena of nonlinear systems may occur when the system is excited with input of frequency ω. A linear system would respond only with the same frequency, but a nonlinear system may experience responses, besides ω, at frequencies ω/n where n is an integer. These are called subharmonic oscillations. Superharmonic oscillations at frequencies $n \cdot \omega$ are also possible although they are not as severe as the subharmonics. This is because higher frequencies are usually associated with more damping. The generation of the above oscillations depends upon the initial conditions, as well as the amplitude and frequency of the excitation.

One question that one may ask is, how many types are behavior are possible for nonlinear systems? The answer to this depends mainly on the system dimensionality. Suppose we have a first order, scalar, system. This involves one variable only, and this can be represented on a straight line. Since it is restricted to move on this line, the system can only experience one or more equilibrium points, and these can be either stable or unstable. Now consider a second order system, this involves two variables x_1, x_2, and if we want to plot these together we need to use a two–dimensional graph, a plane. The solutions in time on this plane can do whatever they desire except cross each other: this would violate uniqueness of solutions for all subsequent times, since different response would be obtained from identical starting conditions. Solutions of dynamic systems, linear or nonlinear, exist and are unique. We can see that two types of behavior are possible here: the solutions can either approach a point asymptotically (equilibrium point), or a closed curve on which they may be constrained to move for ever. This represents a periodic solution. Such an isolated periodic solution is called a *limit cycle* and occurs without any periodic excitation! The study of limit cycles is a very tough but nice problem in nonlinear systems. Now let's imagine a system with three or more state variables. We need at least a three–dimensional graph to plot all of our solutions together here. It is clear that such a system may exhibit both equilibrium points and isolated periodic solutions or limit cycles. In three or more dimensions, the restriction that trajectories may not cross does not constrain the solutions to be simple. There is enough room in three dimensional spaces and beyond so that the solutions they can wrap around each other, twist, turn, and tangle themselves into fantastic knots as they develop in time, forming complicated patterns. Therefore, some complex dynamic behavior is possible for third or higher order systems. Forced and/or discrete systems are usually more complicated. To summarize we can have the following possible types of behavior for nonlinear systems:

- First order unforced systems: Equilibrium points only.
- Second order unforced systems: Equilibrium points and limit cycles.
- Third order or higher unforced systems: Equilibrium points, limit cycles, possible complicated behavior.
- Second order or higher forced systems: Equilibrium points, periodic solutions, possible complicated behavior.
- Discrete systems of any order: Equilibrium points, periodic solutions, possible complicated behavior.

Let's consider as an example, a Van der Pol equation; a spring–mass–damper system with nonlinear damping and no forcing,

$$m\ddot{x} - b(1 - x^2)\dot{x} + kx = 0 \; .$$

The equilibrium point of this equation is $x = 0$, the origin. By linearization we can easily see that the origin is unstable. The linearized system is $m\ddot{x} - b\dot{x} + kx = 0$, and we see that $x = 0$ is unstable because of the negative damping term $-b$. So where are the solutions going? We have seen that for small x the solutions move away from $x = 0$. For large x we can see that the term $-b(1 - x^2)$ will become positive, so the damping will be positive and the solutions will have to move towards $x = 0$. Therefore, solutions which originate from

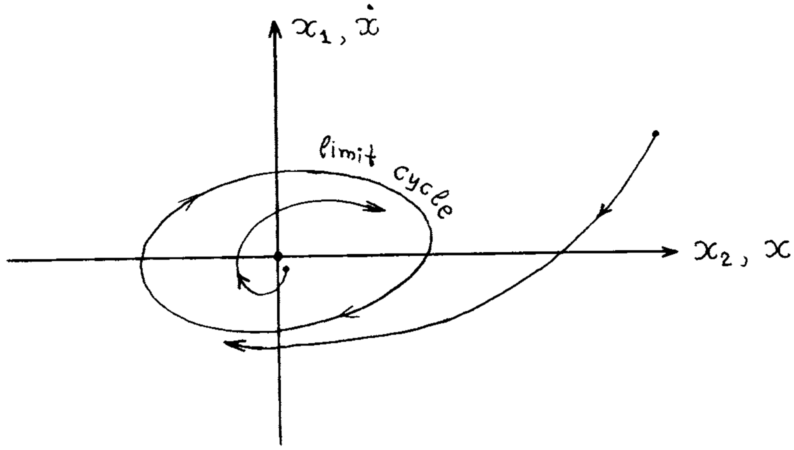

Figure 43: Poincaré's limit cycle prediction technique

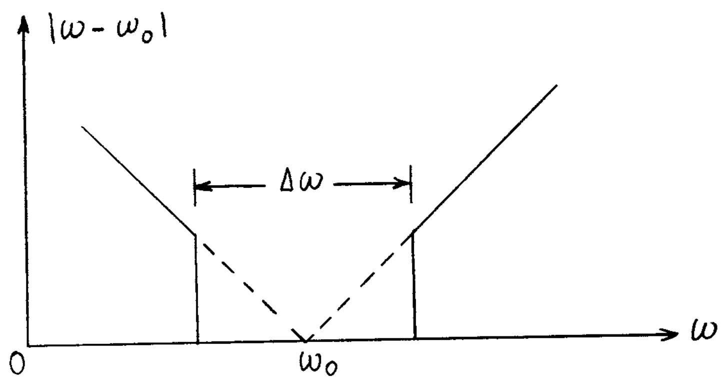

Figure 44: Frequency entrainment

large x will move towards the origin. Since they cannot cross each other and there are no other equilibrium points to attract them, they have to approach a limit cycle which should be located somewhere around the origin. This argument, which is known as the Poincaré–Bendixon theorem, holds for second order systems only and it will reveal the existence of a limit cycle but it cannot provide any information about its size or frequency. The sketch of Figure 43 illustrates Poincaré's argument.

Another phenomenon typical in nonlinear systems is the frequency entrainment. Suppose we have a system which is capable of exhibiting a limit cycle of frequency ω_0. If a periodic force of frequency ω is applied to this system we have the phenomenon of beats. As the difference between the two frequencies decreases, the beat frequency also decreases and, for a linear system, it is zero only if $\omega = \omega_0$. In a self excited nonlinear system, however, it is found that the frequency ω_0 of the limit cycle falls in synchronization with, in other words it is entrained by, the forcing frequency ω within a certain band of frequencies. This phenomenon is illustrated in Figure 44.

8.2 A Simple Zero Eigenvalue

Suppose we have the nonlinear system of state equations,

$$\dot{x} = f(x) \ .$$

We know that the equilibrium points, $\overline{x}$, of the system are defined by

$$f(\overline{x}) = 0 \ .$$

This is a nonlinear system of algebraic equations and it may have multiple solutions in $\overline{x}$, which means that the nonlinear system may have more that one positions of static equilibrium. If we pick one equilibrium, $\overline{x}$, we can establish its stability properties by linearization. The linearized system becomes

$$\dot{x} = Ax \ ,$$

where A is the Jacobian matrix of $f(x)$ evaluated at $\overline{x}$,

$$A = \left.\frac{\partial f}{\partial x}\right|_{\overline{x}} \ ,$$

and the state x has been redefined to designate small deviations from the equilibrium $\overline{x}$,

$$x \rightarrow x - \overline{x} \ .$$

As long as all eigenvalues of A have negative real parts, we know that the linear system will be stable. This means that the equilibrium $\overline{x}$ will be stable for the nonlinear system as well. No surprises so far, in fact what we have just said is nothing but Lyapunov's linearization technique.

The question we ask ourselves next is, what happens if one real eigenvalue of the linearized matrix A is zero? The interesting case here is when the rest of the eigenvalues have all negative real parts, otherwise $\overline{x}$ is unstable and the problem is solved. If the case of a zero eigenvalue appears to be too specialized to be of any practical use consider this: Assume that $f(x)$ depends on one physical parameter (and there will be plenty of physical parameters in any problem) and that physical parameter is allowed to vary over some range; aren't they all? Then it is clear that A will depend on that parameter and as the parameter varies, it is possible that one real eigenvalue of A will become zero for a specific value of the parameter. Our problem is then to establish the dynamics of the nonlinear system as *one real* eigenvalue of A crosses zero; i.e., goes from negative to positive. As the solutions evolve it time, things are interesting only along the direction of the eigenvector that corresponds to the critical eigenvalue (the one that crosses zero). Along the rest of the directions in the state space, everything should converge back to the equilibrium; remember that we assumed that all remaining eigenvalues of A have negative real parts. The above statement should be clear for those of us who haven't forgotten our ME 2801 or O.D.E. material. Although, strictly speaking, it is a true statement for linear systems, there are technical reasons that force it to be true for nonlinear systems as well, the only difference is that the corresponding directions in the state space are curved instead of straight.

We can see then that it is possible to approximate our original system by a one–dimensional system, which is much easier to analyze. The dynamics of the two systems will be qualitatively similar. The formalization of the above reduction procedure consitutes what is known as center manifold reduction, or normal form computation in nonlinear analysis. So let's see what happens for the case of a zero eigenvalue by using a (typical) first order system,

$$\dot{x} = \lambda x - x^3 \ ,$$

where x is scalar and λ is our distinguished parameter which is allowed to vary between -1 and $+1$. The equilibrium points of the system can be found from

$$\lambda x - x^3 = 0 \Longrightarrow x(\lambda - x^2) = 0 \ ,$$

and we can see that, depending on the sign of λ the equilibria are

$$\overline{x} = 0 \ ,$$

if $\lambda < 0$, and

$$\overline{x} = 0 \quad \text{and} \quad \overline{x} = \pm\sqrt{\lambda} \ ,$$

if $\lambda > 0$. There is only one equilibrium for negative λ, this is $\overline{x} = 0$, the trivial equilibrium. However, as λ crosses zero moving towards positive values a new pair of equilibria appears out of thin air. These two new equilibria are symmetric (equal plus and minus values), they are close to the trivial equilibrium initially, but as λ moves away from its critical value, $\lambda = 0$, they move further away from zero. To analyze the stability properties of these equilibria, let's pick $\overline{x} = 0$ first. The Jacobian is,

$$\left.\frac{\partial f}{\partial x}\right|_{\overline{x}} = \lambda - 3\overline{x}^2 \ .$$

At $\overline{x} = 0$ we get the linearized system

$$\dot{x} = \lambda x \ ,$$

and we see that $\overline{x} = 0$ is stable if $\lambda < 0$ and unstable if $\lambda > 0$. For $\overline{x} = +\sqrt{\lambda}$ we get the linearized system

$$\dot{x} = \left[\lambda - 3\left(\sqrt{\lambda}\right)^2\right] x = -2\lambda x \ .$$

We can see then that for $\lambda > 0$, the equilibrium $\overline{x} = +\sqrt{\lambda}$ is stable. Remember that for $\lambda < 0$ this equilibrium does not exist. The same is true for the other equilibrium $\overline{x} = -\sqrt{\lambda}$. Therefore, we can summarize our findings as follows:

- For $\lambda < 0$ only the trivial equilibrium exists and is stable.
- For $\lambda > 0$ the trivial equilibrium becomes unstable and a pair of symmetric stable equilibria are generated.

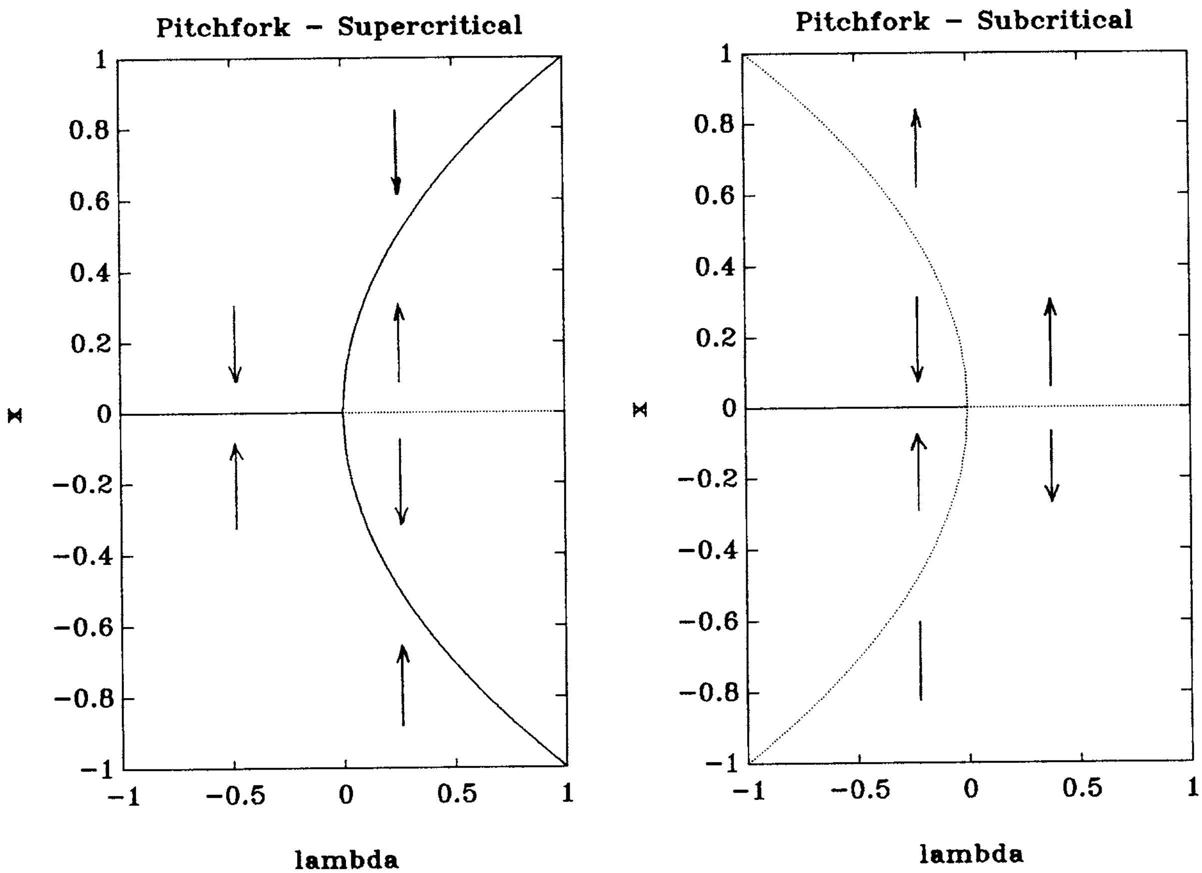

Figure 45: Pitchfork bifurcations

This phenomenon, the loss of stability of an equilibrium and the generation of additional equilibrium states, is called a pitchfork bifurcation and is very common in nature; Euler buckling of a beam is a very typical example. In particular, we refer to the above case as the supercritical pitchfork, this is a rather benign loss of stability since upon loss of stability of the trivial equilibrium the additional nearby equilibrium states are stable. Graphically, we can represent this case as shown in Figure 45 where solid curves represent stable and dotted curves unstable equilibria. We have also indicated the direction of solutions in time of our system for different values of λ. Occasionally, the above case is referred to as a soft loss of stability since for small values of λ beyond its critical value, the final steady state of the system does not differ much from the nominal (trivial) steady state.

As a second example, consider a "similar" system as before, the linear part remains the same, and the nonlinear part x^3 suffers a sign change,

$$\dot{x} = \lambda x + x^3 \ .$$

We can analyze this in exactly the same way as before, and we can draw the following conclusions (verify these),

- For $\lambda > 0$ only the trivial equilibrium exists and is unstable.
- For $\lambda < 0$ the trivial equilibrium becomes stable and a pair of symmetric unstable equilibria are generated.

This case, which is also shown in Figure 45, is called a subcritical pitchfork. A comparison with the previous case reveals that this is a much more serious loss of stability case. Upon loss of stability of the trivial equilibrium position, there is no other stable equilibrium in its vicinity to attract the solutions, which may therefore assume a different state of motion with what could be observed as a discontinuous jump. Furthermore, even before the trivial equilibrium loses its stability the domain of attraction becomes very small and a random perturbation can always throw the system to a different state of motion. This new steady state may be a limit cycle or, depending on the dimensionality of the system, a more complicated response pattern. This loss of stability, sometimes called a hard loss of stability, demonstrates the significance of nonlinear terms in the equations of motion.

8.3 A Purely Imaginary Pair of Eigenvalues

Assume now that our nonlinear system has one pair of purely imaginary eigenvalues for some value of the parameter λ. In other words, this means that as λ is varied over some range, one pair of complex conjugate eigenvalues of the linearized system matrix A crosses the imaginary axis. It is assumed that the rest of the eigenvalues of A remain negative or have negative real parts. We wish to investigate what happens to the nonlinear system during this process. More specifically, in the previous section we saw that the case of one real eigenvalue crossing zero is associated with the generation or exchange of stability of additional equilibrium points for the nonlinear system. The purpose of this section is to show that the corresponding case of the real part of one complex conjugate pair of eigenvalues crossing zero is associated with the generation of periodic solutions or limit cycles for the nonlinear system.

Following similar arguments as before, we can convince ourselves that in the case of a purely imaginary pair of eigenvalues, the only interesting dynamics of $\dot{x} = f(x)$ will be concentrated on a two dimensional space spanned by the eigenvectors which correspond to the critical pair of eigenvalues of A. We start, therefore, with a two dimensional system in the rather special form,

$$\begin{aligned}\dot{x}_1 &= \lambda x_1 - \omega x_2 + a x_1(x_1^2 + x_2^2) \ , \\ \dot{x}_2 &= \omega x_1 + \lambda x_2 + a x_2(x_1^2 + x_2^2) \ .\end{aligned}$$

The system admits the trivial equilibrium $\overline{x}_1 = \overline{x}_2 = 0$. The linearized equations around the trivial equlibrium are

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \lambda & -\omega \\ \omega & \lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} ,$$

with eigenvalues $\lambda \pm i\omega$. Therefore, for $\lambda = 0$ the eigenvalues are purely imaginary (we assume that $\omega \neq 0$). As λ crosses zero, the trivial equilibrium becomes unstable. To compute other potential equilibrium points for our nonlinear system we use

$$\begin{aligned}\lambda\overline{x}_1 - \omega\overline{x}_2 + a\overline{x}_1(\overline{x}_1^2 + \overline{x}_2^2) &= 0 \ , \\ \omega\overline{x}_1 + \lambda\overline{x}_2 + a\overline{x}_2(\overline{x}_1^2 + \overline{x}_2^2) &= 0 \ .\end{aligned}$$

If we multiply the first equation by $\overline{x}_2$, the second by $\overline{x}_1$, and we add them up, we get

$$\omega(\overline{x}_1^2 + \overline{x}_2^2) = 0 \ .$$

Therefore, since $\omega \neq 0$, the only equilibrium solution is the trivial equilibrium $\overline{x}_1 = \overline{x}_2 = 0$. To proceed with the analysis we introduce polar coordinates, (r, θ), by using the transformation,

$$\begin{aligned}x_1 &= r\cos\theta \ , \\ x_2 &= r\sin\theta \ .\end{aligned}$$

The equations of motion are then written as

$$\begin{aligned}\dot{x}_1 = \dot{r}\cos\theta - r\dot{\theta}\sin\theta = \lambda r\cos\theta - \omega r\sin\theta + ar^3\cos\theta \ , \\ \dot{x}_2 = \dot{r}\sin\theta + r\dot{\theta}\cos\theta = \omega r\cos\theta + \lambda r\sin\theta + ar^3\sin\theta \ ,\end{aligned}$$

which reduce to

$$\begin{aligned}\dot{r} &= \lambda r + ar^3 \ , \\ \dot{\theta} &= \omega r \ .\end{aligned}$$

It is clear that an equilibrium point, $\overline{r}$, of the $\dot{r}$ equation will correspond to a limit cycle back in the original coordinates x_1 and x_2. We can see that the $\dot{r}$ equation has equilibria given by

$$\lambda\overline{r} + a\overline{r}^3 = 0 \ .$$

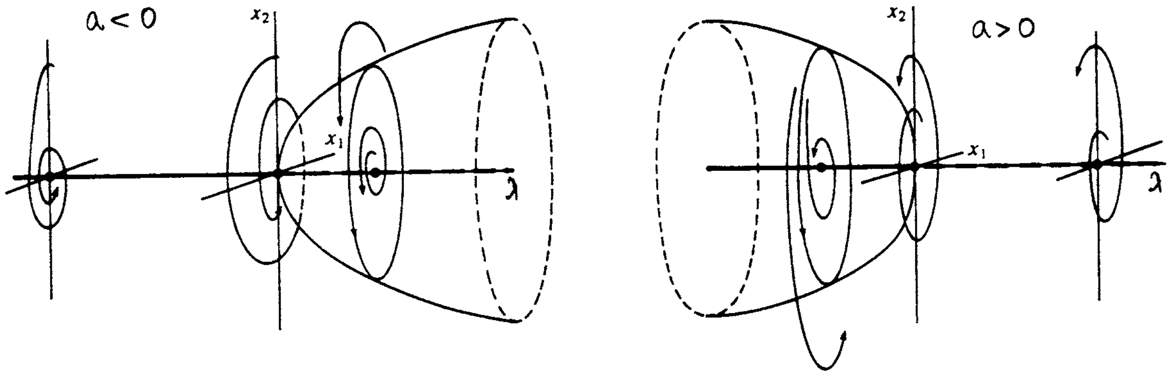

Figure 46: Hopf bifurcations

Let us assume that $a < 0$. Then for $\lambda < 0$, the trivial equilibrium is stable. For $\lambda > 0$ there is a stable limit cycle of radius proportional to the square root of λ surrounding the unstable trivial equilibrium. If $a > 0$, then the limit cycle occurs for $\lambda < 0$; it is unstable and surrounds a stable equilibrium point. The two cases are shown schematically in Figure 46. This resembles our pitchfork bifurcation of the previous section. Therefore, we can summarize our conclusions about the x_1, x_2 system as follows:

- If $a < 0$, then:
  - If $\lambda < 0$ the trivial equilibrium is stable.
  - If $\lambda > 0$ the trivial equilibrium is unstable, and a family of stable limit cycles with amplitude $\pm\sqrt{-\lambda/a}$ exists.
- If $a > 0$, then:
  - If $\lambda > 0$ the trivial equilibrium is unstable.
  - If $\lambda < 0$ the trivial equilibrium is stable, and a family of unstable limit cycles with amplitude $\pm\sqrt{-\lambda/a}$ exists.

We can see that the situation is similar to our pitchfork case; here we have the generation of periodic solutions except of equilibrium points. This bifurcation to periodic solutions is normally called the Poincaré–Andronov–Hopf bifurcation. Analogously to the pitchfork case, we distinguish here the two major cases, supercritical and subcritical Hopf bifurcation. For more complicated systems, the reduction to the above two dimensional form and the computation of the leading nonlinear coefficient a which dictates limit cycle stability can be a significant undertaking.

8.4 Popov and Circle Criteria

Quite often, we need to analyze a control loop which contains a nonlinearity. Such a typical loop is shown in Figure 47. The two methods that we describe here enclose the nonlinearity

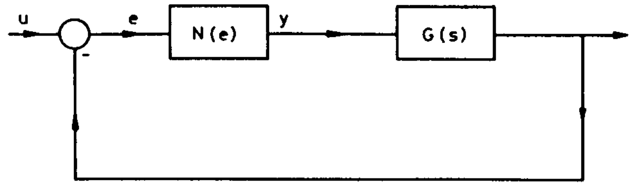

Figure 47: The feedback loop to be analyzed by the Popov or circle method

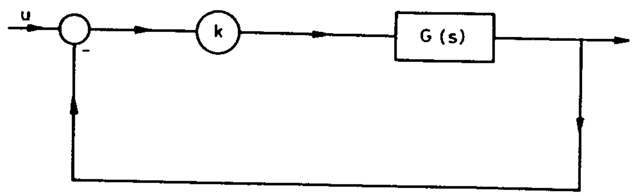

Figure 48: The linear feedback loop that is the subject of Aizerman's and Kalman's conjectures

in a linear envelope. The linear envelope rather than the particular nonlinearity is then used in the subsequent analysis. This approach leads to sufficient but not necessary stability conditions. Before proceeding to describe graphical techniques for the analysis of a feedback loop containing a nonlinearity, it is instructive to consider two celebrated conjectures, by two of the best minds of control theory.

1. The Aizerman and Kalman conjectures:
Aizerman postulated that the system of Figure 47 will be stable provided that the linear system of Figure 48 is stable for all values of k in the interval $[k_1, k_2]$ where k_1, k_2 are defined by the relation

$$k_1 \leq \frac{N(e)}{e} \leq k_2 ,$$

for all $e \neq 0$. In this notation k_1, k_2 represent a linear envelope surrounding the nonlinearity, see Figure 51 where A stands for k_1 and B for k_2. Aizerman's conjecture, reasonable as it might sound, is false as has been shown by counter–examples.

Kalman suggested that the system of Figure 47 will be stable provided that the linear system of Figure 48 is stable for all k in the interval $[\hat{k}_1, \hat{k}_2]$ where

$$\hat{k}_1 \leq \frac{dN(e)}{de} \leq \hat{k}_2 ,$$

and where

$$k_1 \leq \frac{N(e)}{e} \leq k_2 ,$$

and

$$\hat{k}_1 \leq k_1 \leq k_2 \leq \hat{k}_2 .$$

Kalman's conjecture imposes additional requirements on the nonlinear characteristics but nevertheless it is also false — again shown by counter–examples. The failure of the two

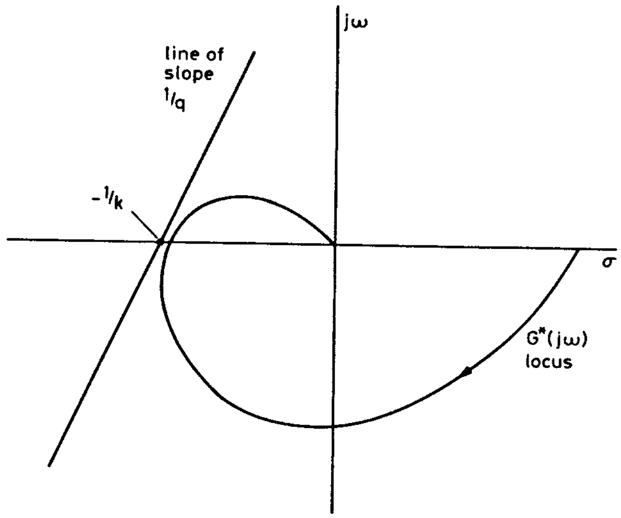

Figure 49: Popov's stability test: The control loop is guaranteed stable

conjectures shows that intuitive reasoning cannot be relied on in nonlinear systems. One reason for the failure of the conjectures is that instabilities may arise in nonlinear systems due to the effects of harmonics. These are, of course, absent in linear systems. In the following, we discuss briefly two techniques for dealing with the problem of Figure 47. These two techniques, Popov's and circle criteria, can be viewed as extensions to Nyquist's stability criterion for linear systems.

2. Popov's stability criterion:
Popov developed a graphical Nyquist–like criterion to examine the stability of the loop shown in Figure 47. It is assumed that $G(s)$ is a stable transfer function. The nonlinearity $N(e)$ must be time–invariant and piecewise continuous function of e. The derivative $dN(e)/de$ must be bounded and $N(e)$ must satisfy the condition

$$0 < \frac{N(e)}{e} < k \ ,$$

for some positive constant k. Graphically, the last condition means that the curve representing N must lie within a particular linear envelope. A *sufficient* condition for *global* asymptotic stability of the feedback loop may then be stated as:

> If there exists any real number q and an arbitrarily small number $\delta > 0$ such that
>
> $$\Re\{(1 + j\omega q)G(j\omega)\} + \frac{1}{k} \geq \delta > 0 \ ,$$
>
> for all ω then for any initial state the system output tends to zero as $t \to \infty$.

The proof can be found in most textbooks on nonlinear control and it makes use of Lyapunov's direct method.

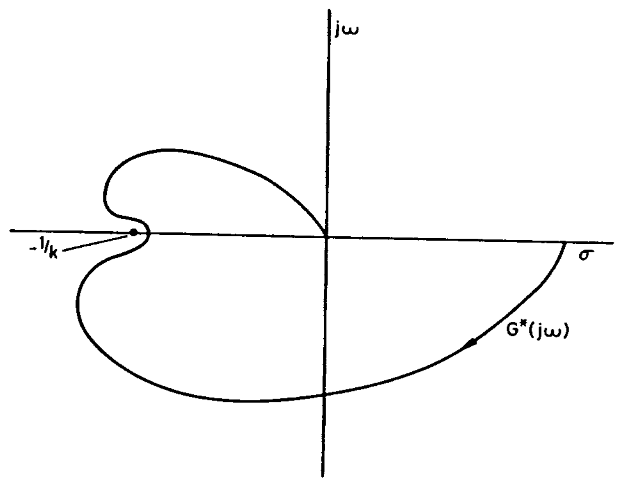

Figure 50: Popov's stability test: No line through the $-1/k$ point avoids intersection with the $G^*(j\omega)$ locus and the loop may be unstable

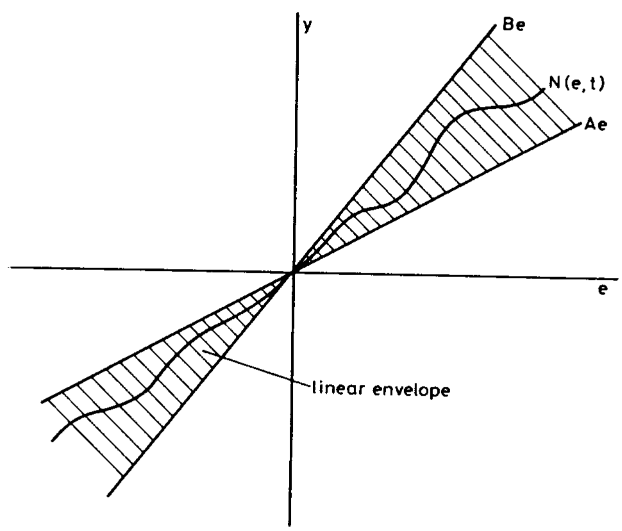

Figure 51: The linear envelope for the circle method

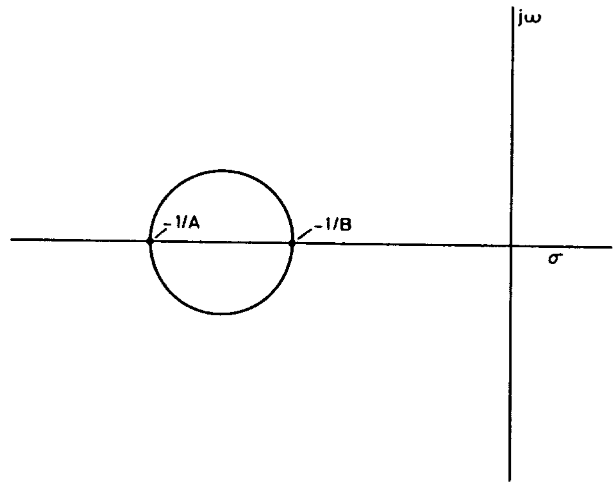

Figure 52: The circle criterion

To carry out a graphical test based on the above equation, a modified transfer function $G^*(j\omega)$ is defined by

$$G^*(j\omega) = \Re\{G(j\omega)\} + j\omega\Im\{G(j\omega)\} \equiv X(j\omega) + jY(j\omega) \ .$$

The criterion then, in terms of X and Y, becomes

$$X(j\omega) - qY(j\omega) + \frac{1}{k} \geq \delta > 0 \ .$$

The $G^*(j\omega)$ curve (the so called Popov locus) is plotted in the complex plane. The system is then stable if some straight line, at an arbitrary slope $1/q$, and passing through the $-1/k$ point avoids intersecting the $G^*(j\omega)$ locus. Figures 49 and 50 show two possible graphical results for stable and not necessarily stable situations respectively. Recall that the test gives a sufficient condition for stability and that the feedback loop whose result is given in Figure 50 is not necessarily unstable.

3. The circle method:
The circle method of stability analysis can be considered as a generalization of Popov's method. Compared with that method it has two important advantages:

1. It allows $G(s)$ to be open loop unstable;
2. It allows the nonlinearity to be time varying.

The nonlinearity N is assumed to lie within an envelope such that,

$$Ae < N(e,t) < Be \ ,$$

as shown in Figure 51. Then it is a sufficient condition for asymptotic stability that the Nyquist plot $G(j\omega)$ lies outside a circle in the complex plane that crosses the real axis at the points $-1/A$ and $-1/B$ and has its center at the point,

$$\frac{1}{2}\left[-\left(\frac{1}{A}+\frac{1}{B}\right)+\frac{1}{2}j\omega q\left(\frac{1}{A}-\frac{1}{B}\right)\right] ,$$

for some real value of q. Here it is assumed that $A < B$.

This is the so called generalized circle criterion. Notice that the center of the circle depends on both frequency and choice of the value of q. In return for a loss of sharpness in the result (remembering that the method gives a sufficient criterion), q can be set equal to zero and then a single frequency invariant circle results (Figure 52). The circle can be considered as the generalization of the $(-1, 0)$ point in the Nyquist test for linear systems.

8.5 Describing Function Analysis

1. Describing Functions

There are a few tools that can be used to predict the existence, magnitude, and stability of limit cycles, namely,

- numerical integrations,
- continuation methods,
- perturbation methods,
- describing function analysis.

Numerical integrations are easy to apply but the can only be used to confirm rather than predict possible behavior, especially when a large number of variables and initial conditions are present. Continuation methods require some initial approximation of the limit cycle for a given set of parameters, while perturbation methods are best applied to system with smooth nonlinearities. Describing function analysis is an approximate method that is best suited to the discontinuous nonlinearities common in several control systems.

Suppose that the input to a nonlinear element is sinusoidal. The output will be periodic and suppose that only the component with the same frequency as the input (the fundamental harmonic component) is significant. The complex quantity

$$G_d = \frac{C_1}{M}\langle\phi_1\rangle ,$$

where

$$\begin{aligned} M &= \text{amplitude of input sinusoid} \\ C_1 &= \text{amplitude of fundamental harmonic component of output} \\ \phi_1 &= \text{phase shift of fundamental harmonic component of output} \end{aligned}$$

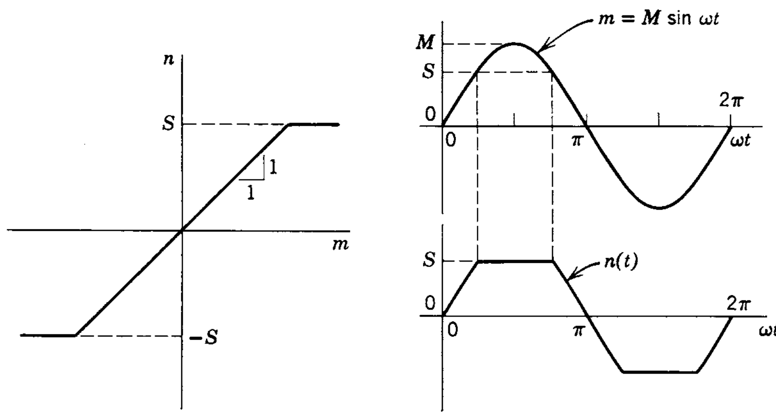

Figure 53: Saturation nonlinearity

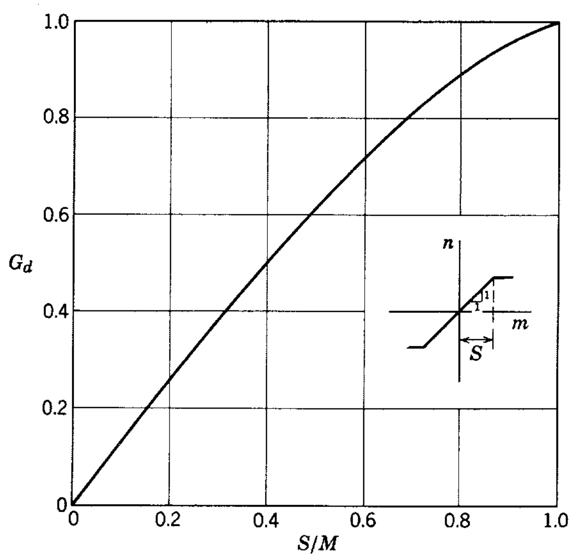

Figure 54: Describing function for saturation

is called the describing function G_d.

2. Computation of Describing Functions

For a sinusoidal input

$$m(t) = M \sin \omega t$$

to the nonlinear element, the output $c(t)$ may be expressed in Fourier series as follows:

$$\begin{aligned} c(t) &= A_0 + \sum_{n=1}^{\infty} (A_n \cos n\omega t + B_n \sin n\omega t) \\ &= A_0 + \sum_{n=1}^{\infty} (C_n \sin(n\omega t + \phi_n) \ , \end{aligned}$$

where

$$\begin{aligned} A_n &= \frac{1}{\pi} \int_0^{2\pi} c(t) \cos(n\omega t)\, d(\omega t) \ , \\ B_n &= \frac{1}{\pi} \int_0^{2\pi} c(t) \sin(n\omega t)\, d(\omega t) \ , \\ C_n &= \sqrt{A_n^2 + B_n^2} \ , \\ \phi_n &= \tan^{-1}\left(\frac{A_n}{B_n}\right) \ . \end{aligned}$$

If the nonlinearity is symmetric, then $A_0 = 0$. The fundamental harmonic component of the output is

$$\begin{aligned} c_1(t) &= A_1 \cos \omega t + B_1 \sin \omega t \\ &= C_1 \sin(\omega t + \phi_1) \ . \end{aligned}$$

The describing function is then given by,

$$G_d = \frac{C_1}{M} \langle \phi_1 \rangle = \frac{\sqrt{A_1^2 + B_1^2}}{M} \left\langle \tan^{-1}\left(\frac{A_1}{B_1}\right) \right\rangle \ .$$

As an example, consider the saturation nonlinearity of Figure 53. A Fourier calculation of the output waveform for a sinusoidal input gives the following describing function

$$G_d = \frac{2}{\pi} \left[\sin^{-1}\left(\frac{S}{M}\right) + \frac{S}{M} \sqrt{1 - \left(\frac{S}{M}\right)^2} \right] \ .$$

For a stauration function of slope k the term $2/\pi$ in front of the above expression becomes $2k/\pi$. Also, this expression is true for $S < M$. For $S > M$, the input signal does not feel the effcts of the saturation and it behaves just like a linear unity gain; i.e., $G_d = 1$ for $S > M$. A plot of the saturation describing function G_d versus the dimensionless ratio S/M is shown in Figure 54. A very useful general property for calculating describing functions is:

> The describing function of the sum of two elements is the sum of the individual describing functions.

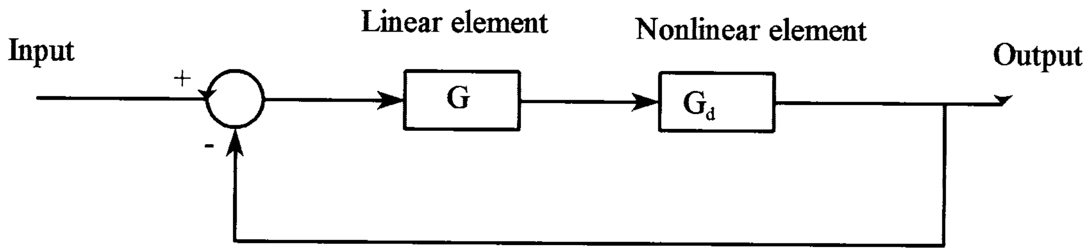

Figure 55: Describing function analysis

3. Describing Function Analysis

Consider the closed–loop feedback system of Figure 55 containing a linear element with transfer function G and a nonlinear element with describing function G_d. If the higher harmonics are sufficiently attenuated, the describing function G_d can be treated as a complex gain. Then, the closed loop frequency response is

$$\frac{C(j\omega)}{R(j\omega)} = \frac{G_d G(j\omega)}{1 + G_d G(j\omega)} \ .$$

The characteristic equation is

$$1 + G_d G(j\omega) = 0 \ ,$$

or

$$G(j\omega) = -\frac{1}{G_d(M)} \ .$$

If this equation is satisfied, then the system will exhibit a limit cycle with frequency ω and amplitude M found from the intersection of $G(j\omega)$ and $-1/G_d(M)$ graphs.

4. Stability of Limit Cycles

To assess the stability of these limit cycles, we have to recognize the similarity between the above and the Nyquist criterion for linear systems. For example, consider the case shown in Figure 56. We see that we have two limit cycles with characteristics (M_A, ω_A) and (M_B, ω_B) with $M_A < M_B$. Consider the intersection A of the $G(j\omega)$ and $-1/G_d(M_A)$ loci and assume a small decrease in amplitude M_A. The representative point on the $-1/G_d$ locus will move to a new point, D. This point is not encircled by the $G(j\omega)$ locus, the system will move further and further away from the intersection and the oscillations will eventually stop. Therefore, point A possesses divergent characteristics and it corresponds to an unstable limit cycle. By a similar argument we can see that point B possesses convergent characteristics and it corresponds to a stable limit cycle. Indeed, if the amplitude of the limit cycle is decreased so that the system moves to point F we can see that the new point is encircled by the $G(j\omega)$ locus, the oscillations will grow, the system will tend to return to the original intersection B and the oscillations are stable. As a summary, we can conclude that in general: The limit cycle is predicted to be stable or unstable according as the locus of $-1/G_d$ crosses the locus of G (the Nyquist plot) from right to left or from left to right, respectively, as M increases,

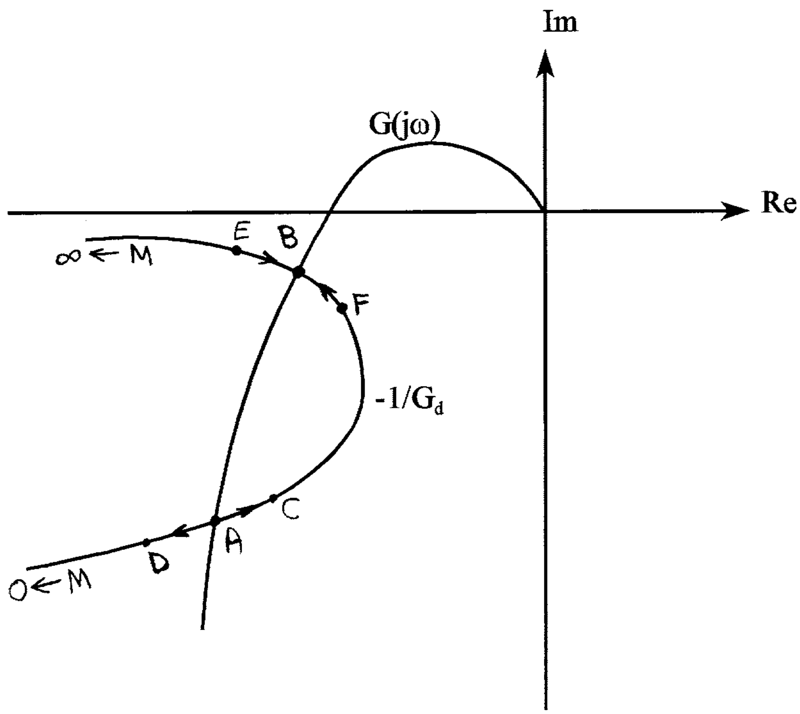

Figure 56: Stability of limit cycles

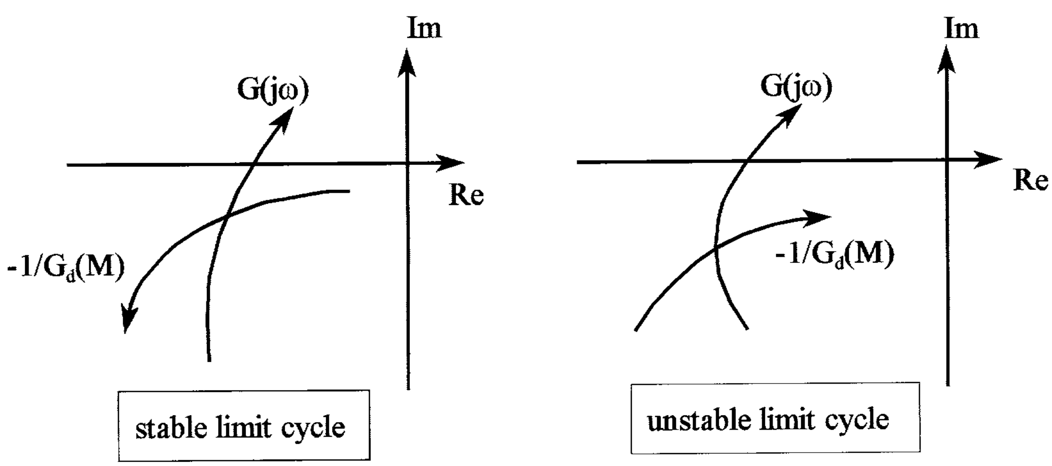

Figure 57: Stable and unstable limit cycles

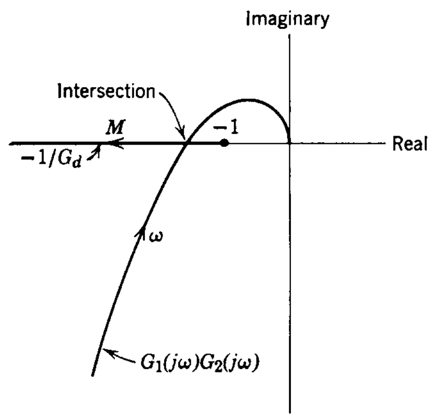

Figure 58: Describing function analysis of saturation in a control loop

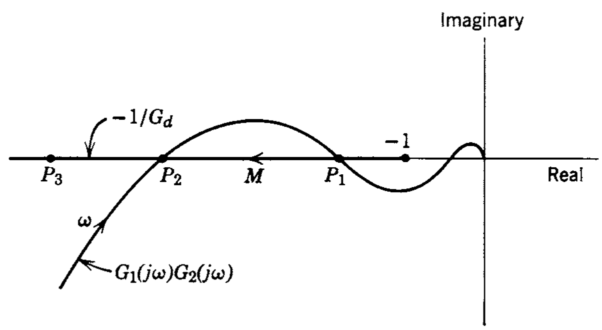

Figure 59: Describing function analysis of saturation in a conditionally stable loop

viewed along the direction of increasing ω. This criterion is illustrated by the sketch of Figure 57.

5. Example: Saturation

Consider a linear system with the saturation nonlinearity shown in Figure 53. Suppose that the Nyquist diagram for the linear element encloses the -1 point, so that the linear system is unstable. If there were no saturation, this means that oscillations with ever–increasing amplitude would develop. To analyze the effect of saturation let us superimpose the graph of the describing function of the saturation nonlinearity onto the Nyquist diagram, as shown in Figure 58. We can see that the effect of the saturation (i.e., limit on actuator stroke) is to generate a stable limit cycle at the intersection point and thus prevent the motions from becoming arbitrarily large. If the gain of the transfer function is decreased so that the locus of $-1/G_d$ does not intersect that of G, the system becomes stable and any oscillations that may develop will eventually die out. No limit cycle (self sustained oscillation) will exist at steady state.

As another example consider the effects of saturation on a conditionally stable system as shown in Figure 59. The linear system is here stable since the polar plot avoids the -1 point. In this case we can see that two limit cycles are created one at P_1 and another one at

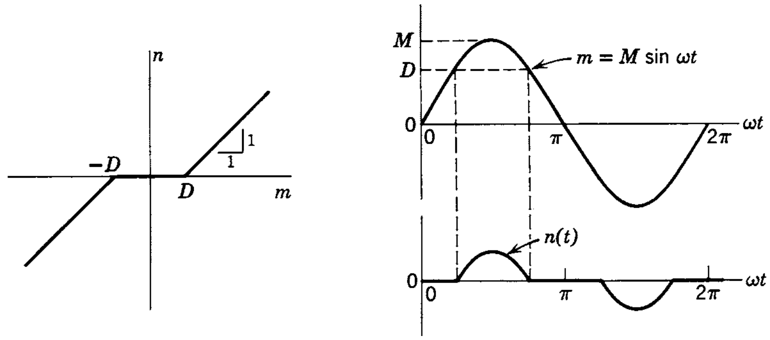

Figure 60: Deadband nonlinearity

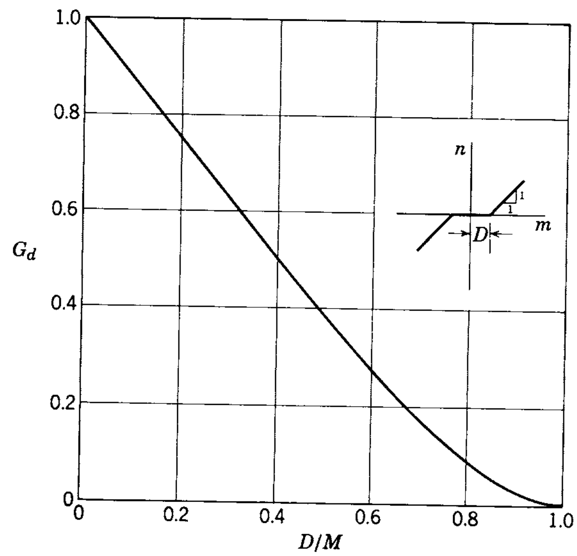

Figure 61: Describing function for deadband

P_2. The limit cycle at P_1 is unstable, whereas the limit cycle at P_2 is stable. Therefore, if the system amplitude exceeds this value, for example during transient response, self–sustained oscillations with amplitude corresponding to P_2 will develop. In this case even though the origin is stable, the effect of the saturation is to limit the origin's domain of attraction. System response will converge to zero as long as the initial transient does not exceed P_1.

6. Example: Deadband

A deadband nonlinearity (Figure 60) can result from Coulomb friction and from overlap of valve ports in hydraulic systems. The linear gain of the deadband is normalized to one and any gain present would be considered as part of the linear portion of the loop. Analysis of the output waveform gives the following describing function

$$G_d = \frac{2}{\pi}\left[\frac{\pi}{2} - \sin^{-1}\left(\frac{D}{M}\right) - \frac{D}{M}\sqrt{1-\left(\frac{D}{M}\right)^2}\,\right] ,$$

which is plotted in Figure 61.

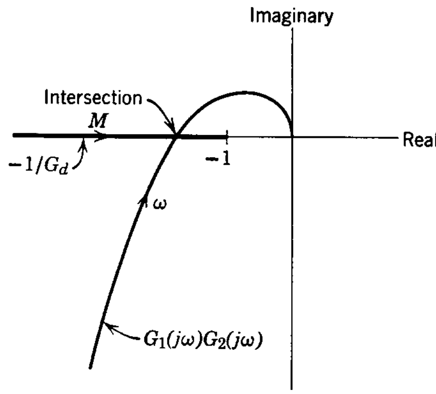

Figure 62: Describing function analysis of deadband in a control loop

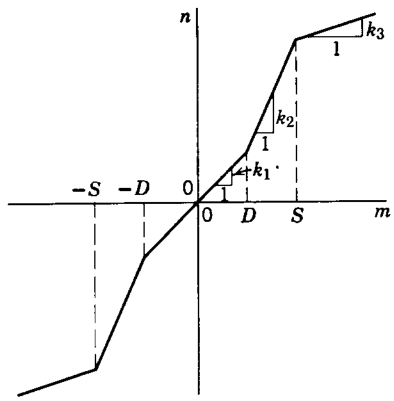

Figure 63: Nonlinear gain characteristic

We note that $-1/G_d$ is a large negative real number for small inputs to the deadband element and approaches -1 for large inputs. Suppose the polar plot is as shown in Figure 62. The linear system with this Nyquist plot would be unstable. The limit cycle at the intersection point is also unstable. This means that the system will actually be stable for small inputs to the deadband (i.e., as long as the intersection point is not crossed over). If it seems peculiar that an unstable linear system may become stable with the addition of a nonlinear element, this is due to the fact that the actual system including the deadband has very small gain at the origin. In this case, since the deadband generates an unstable limit cycle, unbounded oscillations will occur if the input to the deadband is large enough. This is why deadbands are quite undesireble from the stability point of view. In any practical system, however, the deadband will saturate and the oscillations will become bounded. This case is treated next.

7. Example: Nonlinear Gain Characteristics

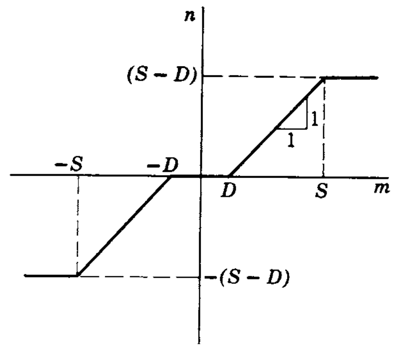

Figure 64: Saturation and deadband nonlinearity

The describing function of the general nonlinear gain characteristic in Figure 63 is,

$$
\begin{aligned}
G_d &= k_3 + \frac{2}{\pi}(k_1 - k_2)\left[\sin^{-1}\left(\frac{D}{M}\right) + \frac{D}{M}\sqrt{1 - \left(\frac{D}{M}\right)^2}\right] \\
&\quad + \frac{2}{\pi}(k_2 - k_3)\left[\sin^{-1}\left(\frac{S}{M}\right) + \frac{S}{M}\sqrt{1 - \left(\frac{S}{M}\right)^2}\right] .
\end{aligned}
$$

The describing functions for saturation and deadband can be obtained from this expression by letting appropriate quantities be zero. With so many parameters involved, it is better to look at a particular case. Of interest is a combination of saturation and deadband (Figure 64). In this case $k_1 = k_3 = 0$ and $k_2 = 1$ and the describing function is plotted in Figure 65. Note that the "gain" is small for small inputs, increases to a maximum, then decreases as the input amplitude M increases. Thus, the quantity $-1/G_d$ starts at $-\infty$ for small inputs, decreases to a minimum, then again approaches $-\infty$ as the input becomes very large. The $-1/G_d$ locus and a polar plot of a linearly unstable system are shown in Figure 66. For the intersections shown, point P_1 is an unstable limit cycle and P_2 is a stable limit cycle. Note that this system is stable for small inputs not exceeding P_1, but once the input amplitude becomes greater than at point P_2, oscillations will build up to a limit cycle at P_2. The $-1/G_d$ locus has a minimum which approaches but never exceeds the -1 point. Thus, a system having this characteristic and designed so that the polar plot does not encircle the -1 point would be stable. However, it is possible for the system to be stable even if the -1 point is encircled because of the minimum of the $-1/G_d$ locus.

8. Backlash and Hysteresis

Backlash and hysteresis nonlinearities are multivalued. With backlash, the input must be moved by a certain amount before any motion of the output occurs. Similarly upon reversal. Generally speaking, backlash can pose a serious threat to the stability of a loop. Dither is a widely used method of removing backlash. Its is very effective where the backlash is caused by friction. Dither is a high frequency signal of constant amplitude and frequency which is added to the control signal at the input to the nonlinearity and has the effect of making the element appear linear. However, dither cannot be used in certain cases such as gear backlash because it is difficult to inject, causes wear, and shows in the output.

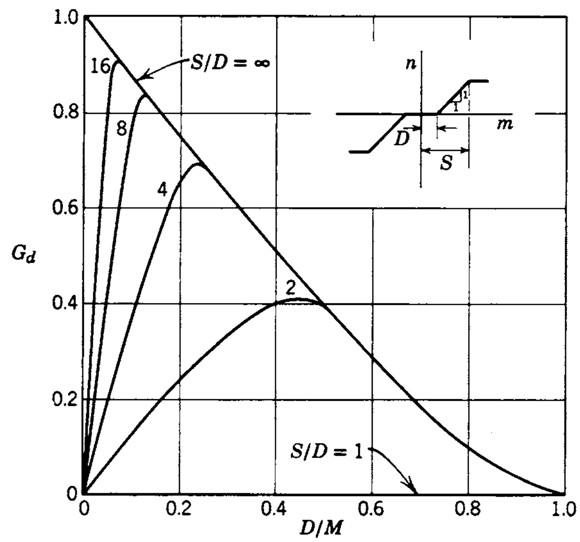

Figure 65: Describing function for saturation and deadband

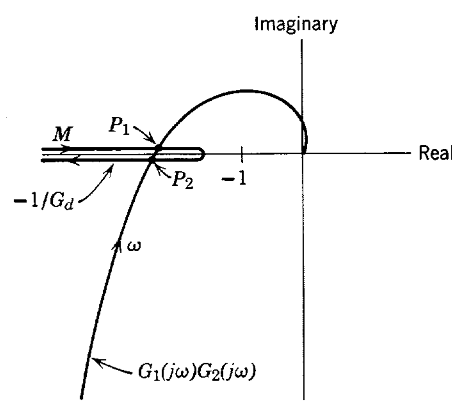

Figure 66: Describing function analysis of saturation and deadband nonlinearity in a control loop

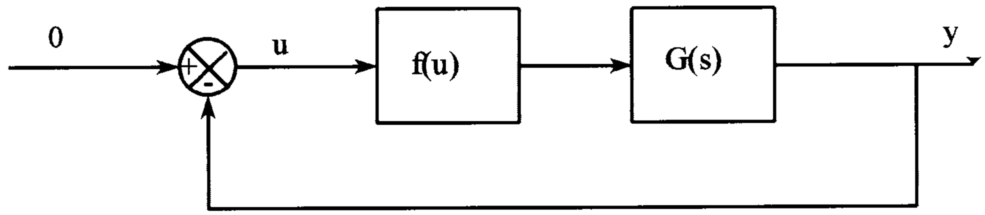

Figure 67: Feedback representation of Van der Pol's equation

Hysteresis nonlinearities constitute a nuisance but not a serious threat to stability. The most noticeable attribute of elements with hysteresis nonlinearity is an amount of phase lag at low frequencies.

9. Comments

The describing function analysis is an extension of linear techniques to the study of nonlinear systems. Typical applications are to systems with few nonlinearities. The analysis is only approximate: there are instances where the describing function analysis predicts the existence of limit cycles but the actual system exhibits none, and other instances where the situation is reversed.

It is more accurate to state that the describing function analysis predicts the likelihood of limit cycles. The system may exhibit a periodic solution with amplitude and frequency close to the predicted ones. Final response has to be verified by numerical integrations.

10. A Counter–example: Van der Pol's Equation

Once more, consider Van der Pol's equation

$$\ddot{y} + \epsilon(3y^2 - 1)\dot{y} + y = 0 \ .$$

In order to represent this in a "block diagram" form including an appropriate nonlinear element, we write it as,

$$\begin{aligned}
&\ddot{y} - \epsilon\dot{y} + y = -3\epsilon y^2\dot{y} \quad \text{or} \\
&\ddot{y} - \epsilon\dot{y} + y = -\epsilon\frac{d}{dt}y^3 \quad \text{or} \\
&(s^2 - \epsilon s + 1)y = \epsilon s(-y^3) \quad \text{or} \\
&\frac{y}{u^3} = \frac{\epsilon s}{s^2 - \epsilon s + 1} \ .
\end{aligned}$$

Therefore, in feedback form,

$$G(s) = \frac{\epsilon s}{s^2 - \epsilon s + 1} \ ,$$

with the nonlinearity $f(u) = u^3$, and zero reference input, so that $u = -y$, see Figure 67. For the cubic nonlinearity,

$$G_d = \frac{3M^2}{4} \ .$$

In order to predict the limit cycle we have to solve

$$G(j\omega) = -\frac{1}{G_d(M)} \ ,$$

or

$$\frac{\epsilon j\omega}{-\omega^2 - \epsilon j\omega + 1} = -\frac{4}{3M^2} \ ,$$

or

$$4(\omega^2 - 1) + j(4 - 3M^2)\epsilon\omega = 0 \ .$$

Therefore, the frequency of the limit cycle is predicted at

$$\omega = 1 \quad (\text{period } 2\pi) \ ,$$

and its amplitude at

$$M = \frac{2}{\sqrt{3}} \ .$$

The graphical construction easily shows that this limit cycle is stable.

Now although Van der Pol's equation cannot be solved analytically, it is possible to obtain asymptotically exact expressions for the limit cycle parameters as ϵ approaches zero or infinity. In the small parameter limit ($\epsilon \to 0$), the equation becomes that of a simple harmonic oscillator with unit angular frequency, coinciding with the prediction of the describing function method. In the large parameter limit ($\epsilon \to \infty$), a perturbation analysis predicts period 1.614ϵ, instead of fixed 2π. In order to understand why the method fails in this case, take a closer look at the frequency response of the linear component:

$$G(j\omega) = \left[-1 + \frac{j}{\epsilon}\left(\omega - \frac{1}{\omega}\right)\right]^{-1} \ .$$

It is clear that, as ϵ increases, so does the range of ω over which $G(j\omega) \approx -1$. This means that in the limit of infinite ϵ we obtain an "all–pass" filter, and hence the harmonic content of the limit cycle becomes such that the predominant response is no longer simply sinusoidal, and the describing function approximation cannot be expected to be valid any more.