

# An Introduction to Nonlinearity in Control Systems

Derek Atherton



Download free books at

**bookboon.com**

Derek Atherton

# An Introduction to Nonlinearity in Control Systems

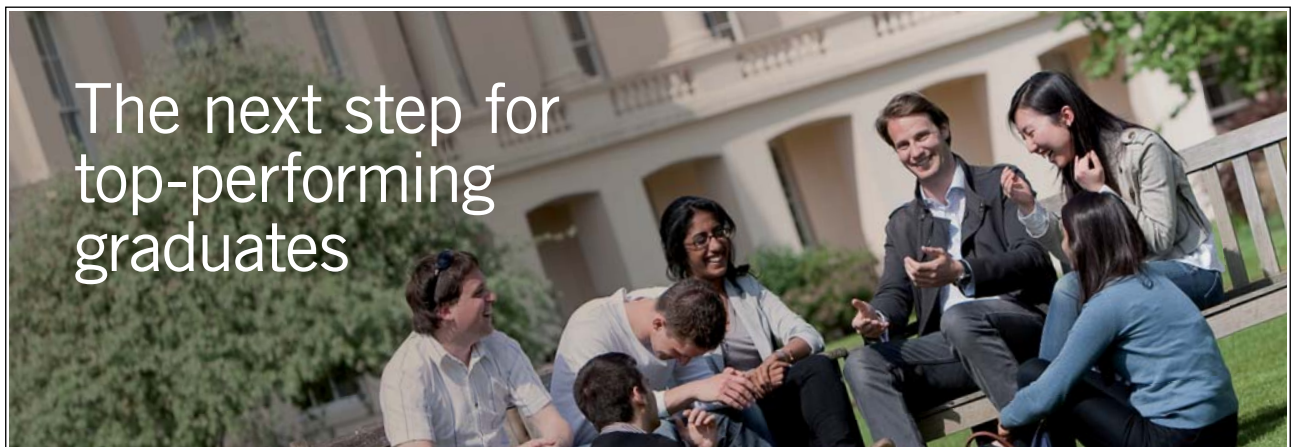
---

An Introduction to Nonlinearity in Control Systems  
© 2011 Derek Atherton & Ventus Publishing ApS  
ISBN 978-87-7681-790-9

# Contents

	<b>Preface</b>	<b>10</b>
<b>1</b>	<b>Introduction</b>	<b>12</b>
1.1	What is nonlinearity?	12
1.2	Forms of nonlinearity	13
1.3	Structure and Behaviour	16
1.4	Overview of contents	17
1.5	References	18
<b>2</b>	<b>The Phase Plane Method</b>	<b>19</b>
2.1	Introduction	19
2.2	Basic Principles	19
2.2.1	The Linear Case	20
2.2.2	The Nonlinear Case	23
2.2.3	An Example	24
2.3	The Phase Plane for Systems with Linear Segmented Nonlinearities	26
2.3.1	Example 1 – Nonlinear Output Derivative Feedback	27
2.3.2	Example 2 – Relay Position Control	28
2.3.3	Example 3 – Position Control with Torque Saturation	34

Please click the advert



## Masters in Management



Designed for high-achieving graduates across all disciplines, London Business School's Masters in Management provides specific and tangible foundations for a successful career in business.

This 12-month, full-time programme is a business qualification with impact. In 2010, our MiM employment rate was 95% within 3 months of graduation\*; the majority of graduates choosing to work in consulting or financial services.

As well as a renowned qualification from a world-class business school, you also gain access to the School's network of more than 34,000 global alumni – a community that offers support and opportunities throughout your career.

For more information visit [www.london.edu/mm](http://www.london.edu/mm), email [mim@london.edu](mailto:mim@london.edu) or give us a call on **+44 (0)20 7000 7573**.

\* Figures taken from London Business School's Masters in Management 2010 employment report

2.4	Conclusions	35
2.5	References	35
2.6	Bibliography	36
<b>3</b>	<b>The Describing Function</b>	<b>37</b>
3.1	Introduction	37
3.2	The Sinusoidal Describing Function	37
3.3	Some Properties of the DF	39
3.4	The Evaluation of some DFs	41
3.4.1	Cubic Nonlinearity	41
3.4.2	Saturation Nonlinearity	41
3.4.3	Relay with Dead Zone and Hysteresis	42
3.5	Nonlinear Models and DFs	44
3.6	Harmonic Outputs	45
3.7	Sine plus Bias DF and the IDF	46
3.8	Conclusions	47
3.9	References	47
3.10	Bibliography	47
3.11	Appendix -Tables of Describing Functions	47
<b>4</b>	<b>Stability and Limit Cycles using the DF</b>	<b>51</b>
4.1	Introduction	51

Please click the advert



You're full of *energy*  
and *ideas*. And that's  
just what we are looking for.

Looking for a career where your ideas could really make a difference? UBS's Graduate Programme and internships are a chance for you to experience for yourself what it's like to be part of a global team that rewards your input and believes in succeeding together.

Wherever you are in your academic career, make your future a part of ours by visiting [www.ubs.com/graduates](http://www.ubs.com/graduates).

[www.ubs.com/graduates](http://www.ubs.com/graduates)



© UBS 2010. All rights reserved.

4.2	Limit Cycle Evaluation	52
4.3	Stability of a Predicted Limit Cycle	53
4.4	DF Accuracy	55
4.5	Some Examples of Limit Cycle Evaluation	55
4.5.1	The Van der Pol Equation	55
4.5.2	Feedback Loop Containing a Relay with Dead Zone	56
4.5.3	Feedback Loop with On Off Relay	58
4.5.4	Use of the IDF	59
4.6	More than one Nonlinear Element	61
4.7	Applications of SBDF to find Limit Cycles	62
4.7.1	Asymmetrical nonlinearity	63
4.7.2	Effect of Bias with Odd Symmetrical Nonlinearity	65
4.8	Conclusions	67
4.9	Bibliography	67
<b>5</b>	<b>The SSDF and Harmonically Forced Systems</b>	<b>68</b>
5.1	Introduction	68
5.2	The Cubic Nonlinearity with Two Sinusoidal Inputs	69
5.3	Modified Nonlinearities and the SSDF	70
5.3.1	Power Law Characteristic	72
5.3.2	Harmonic Nonlinearity	72
5.3.3	Ideal Relay	73

Please click the advert

Discover the truth at [www.deloitte.ca/careers](http://www.deloitte.ca/careers)**Deloitte.**

© Deloitte &amp; Touche LLP and affiliated entities.

5.4	The IDF for Related Signals	74
5.5	More Accurate Determination of Limit Cycles	76
5.6	Closed Loop Frequency Response	79
5.7	Jump Resonance	83
5.7.1	Jump resonance region- use of the IDF	85
5.7.2	Some Examples of Jump Resonance	85
5.8	Conclusions	95
5.9	References	96
<b>6</b>	<b>Limit cycles in relay systems</b>	<b>97</b>
6.1	Introduction	97
6.2	The Frequency Domain Approach	98
6.3	Properties and Evaluation of A loci	101
6.4	Solving for Limit Cycles	102
6.4.1	Relay with no Dead Zone	102
6.4.2	Relay with Dead Zone	104
6.5	Limit Cycle Stability	105
6.6	Some Interesting Limit Cycle Problems	106
6.6.1	Example 1 – Invalid Continuity Condition	106
6.6.2	Example 2 – Multipulse Limit Cycles	107
6.6.3	Example 3 – Limit Cycle with a Sliding Mode	108
6.6.4	Example 4 – Multiple Limit Cycle Solutions	111

Please click the advert

It's only an opportunity if you act on it

[IKEA.SE/STUDENT](http://IKEA.SE/STUDENT)

**IKEA**

© Inter IKEA Systems B.V. 2009



6.6.5	Example 5 – Chaotic Motion	111
6.6.6	Example 6 – Unstable Limit Cycles and Simulation	114
6.7	Forced oscillations	114
6.8	Conclusions	114
6.9	References	114
6.10	Appendix	116
<b>7</b>	<b>Controller Tuning from Relay Produced Limit Cycles</b>	<b>125</b>
7.1	Introduction	125
7.2	Knowledge from the Limit Cycle	127
7.2.1	Relay Autotuning	127
7.2.2	Plant Parameter Identification	129
7.3	Tuning the Controller	130
7.3.1	An Example	133
7.4	Autotuning using the Relay in Parallel with the Controller	133
7.4.1	An Example	136
7.5	Conclusions	137
7.6	References	137
<b>8</b>	<b>Absolute Stability Results</b>	<b>138</b>
8.1	Introduction	138
8.2	Lyapunov's Method	138

Please click the advert

## YOUR CHANCE TO CHANGE THE WORLD

Here at Ericsson we have a deep rooted belief that the innovations we make on a daily basis can have a profound effect on making the world a better place for people, business and society. Join us.

In Germany we are especially looking for graduates as Integration Engineers for

- Radio Access and IP Networks
- IMS and IPTV

We are looking forward to getting your application!  
To apply and for all current job openings please visit our web page: [www.ericsson.com/careers](http://www.ericsson.com/careers)

ericsson.  
com





8.2.1	Definitions of Stability	139
8.2.2	Positive Definite Functions	140
8.2.3	Lyapunov's Theorems	140
8.3	Application of Lyapunov's Method	141
8.3.1	Linear System	141
8.3.2	Nonlinear System	142
8.4	Definitions and Loop Transformations	142
8.4.2	Loop Transformations	143
8.5	Frequency Domain Criteria	144
8.5.1	Popov Theorem	144
8.5.2	The Circle Criterion	145
8.5.3	The Off-axis Circle Criterion	147
8.6	Examples	148
8.7	Conclusions	150
8.8	References	150
8.9	Bibliography	150
<b>9</b>	<b>Design of Nonlinear Control Systems</b>	<b>151</b>
9.1	Introduction	151
9.2	Linearization	152
9.3	Frequency Response Shaping	153
9.4	Nonlinear Compensation	154
9.5	Compensation using DF Models	155
9.6	Time Optimal Control	155
9.7	Sliding Mode Control	157
9.7.1	A Second Order System	157
9.7.2	Further Comments	162
9.8	Fuzzy Logic	163
9.9	Neural Networks	163
9.10	Exact Linearization	164
9.10.1	Relative Degree	166
9.10.2	Input-Output Linearization	167
9.10.3	Exact State Linearization	169
9.10.4	Examples	169
9.10.5	Further Comments	174
9.11	General Comments	175
9.12	References	175

# Preface

The book is intended to provide an introduction to the effects of nonlinear elements in feedback control systems. A central topic is the use of the Describing Function (DF) method since in combination with simulation it provides an excellent approach for the practicing engineer and follows on logically from a first course in classical control, such as the companion volume in this series. Some of the basic material on the topic can be found in my earlier book which is frequently referenced throughout the text.

The first chapter provides an introduction to nonlinearity from the basic definition to a discussion of the possible effects it can have on a system and the different behaviour that might be found, in particular when it occurs within a cascade system of elements or in a feedback loop. The final part of the chapter gives a brief overview of the contents of the book.

Phase plane methods for second order systems are covered in the second chapter. Many systems, particularly electromechanical ones, can be approximated by second order models so the concept can be particularly useful in practice. The second order linear system is first considered as surprisingly it is rarely covered in linear control system texts. The method has the big advantage in that the effects of more than one nonlinear element may be considered. The study is supported by several simulations done in Simulink including one where sliding motion takes place.

The third chapter is the first of three devoted to the study of feedback loops using DF methods. Although the method is an approximate technique its value, and limitations, are supported by a large number of examples containing analytical results and simulations, including the estimation of limit cycles and loop stability. Many early papers on DFs showing how theories could be used to predict specific phenomena were supported by simulations done on analogue computers. Here results from digital simulations using Simulink are presented and this allows much more control of initial conditions to show how different modes may exist dependent on the initial conditions. In particular, Chapter 5 contains some more advanced work including some new results on jump resonance, so some readers may wish to omit this chapter on a first reading.

A relay is a unique nonlinear element in that its output does not depend upon the input at all times but is determined by when the input passes through the relay switching levels. It is this feature which allows the exact determination of limit cycles and their stability in a feedback loop. The basic theory is presented and some simple examples covered. It is then shown in section 6.6 how the approach can be used with computational support to analyse quite complicated periodic modes in relay systems. Among the topics covered for the first time in a textbook are the evaluation of limit cycles with multiple pulses per cycle, as found in a satellite attitude control system, the determination of a limit cycle with sliding and other more advanced aspects which some readers may wish to omit.

A practical method developed in recent years for finding suitable parameters, or tuning, for a controller based on the so called loop cycling method of Ziegler and Nichols has used relay produced limit cycles. Chapter 7 covers these ideas using both approximate analysis based on the DF and exact analysis based on the relay methods of the previous chapter.

Chapter 8 covers the topic of absolute stability namely trying to obtain necessary and sufficient conditions for the stability of a feedback loop with a single nonlinear element. This problem has exercised the minds of theorists for nearly a century but a solution seems no nearer! Several necessary but not sufficient results are presented, which dependent on one's viewpoint may be regarded as 'conservative' or 'robust'. The former is usually the case when one has a mathematically defined nonlinearity and the latter may be used because the result gives stability for any nonlinearity with certain properties, for example, lying within a sector.

The final chapter, chapter 9, discusses quite briefly various methods which can be used for the design of nonlinear systems. The intent has been to provide sufficient information on the methods and their possible advantages and disadvantages. Several of them have complete books written on the topic and more detail could not have been given without the coverage of more specialised mathematics.

Finally my thanks to the University of Sussex for the use of an office and access to the computing facilities during my retirement; to my good friend Keith Godfrey and his student Roland for the computational input on jump resonance and their companionship at the races and to my wife, Constance, for her love and support.

Derek P Atherton  
University of Sussex  
Brighton  
May 2011.

# 1 Introduction

## 1.1 What is nonlinearity?

In order to analyse the behaviour of engineering systems mathematical models are required for the various components. It is common practice to try and obtain linear models as a rich mathematical theory exists for linear systems. These models will always be approximate, although possibly quite accurate for defined ranges of system variables, but inevitably nonlinear effects will eventually be found for large excursions of system variables. 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 circuit theory when an RLC circuit has both d.c and a.c input voltages, the voltage or current elsewhere in the circuit can be found by summing the results of separate analyses for the d.c. and a.c. inputs taken individually, also if the magnitude of the a.c. voltage is doubled then the a.c. voltages and currents elsewhere in the circuit will be doubled.

Thus, mathematically a linear system 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 further point is that when the a.c. voltage is sinusoidal, which is normally assumed in using the term a.c., then all the other voltages and currents in the circuit are sinusoidal with the same frequency as the input.

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

$$y(t) = cx(t) + dx^3(t) \quad (1.1)$$

where the output is the summation of a linear and a nonlinear, cubic, term. This description of nonlinearity is given by a simple mathematical expression. Practical nonlinearities which occur in control engineering, however, can often not be so easily described; so that approximating them may require more complex mathematical models, such as a high order power series or a linear segmented approximation.

More commonly a nonlinear differential equation, for example

$$d^2y(t)/dt^2 + c(dy(t)/dt)^3 + y(t) = x(t) \quad (1.2)$$

will describe the behavior. From an engineering viewpoint it may be desirable to think of this equation in terms of a block diagram, as shown in Figure 1.1, consisting of linear dynamic elements and a static nonlinearity, in this case a cubic, which with input  $dy(t)/dt$  gives an output  $c(dy(t)/dt)^3$ . Unfortunately there is no general approach to solving nonlinear, unlike linear, differential equations. The 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.

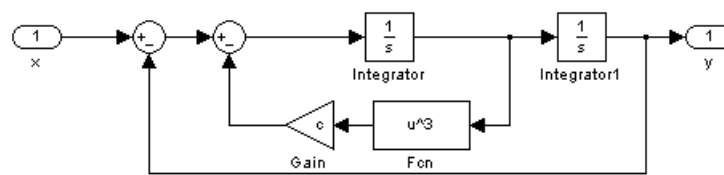


Figure 1.1 Block diagram in Simulink for equation 1.2

A further and very important point, is that unlike a linear operation, a nonlinear operation on a sinusoid of frequency,  $f$ , will not produce an output at frequency  $f$ , alone. For example, if such a sinusoid is applied to the static nonlinearity of equation (1.1) it is easy to show from substituting  $x(t) = a \cos \omega t$  where  $\omega = 2\pi f$  that there are outputs at frequencies  $f$  and  $3f$  of magnitudes  $ca + 3da^3/4$  and  $da^3/4$ , respectively. Perhaps the most interesting aspect of nonlinear systems is that, as will be shown later, they exhibit several forms of unique behaviour which are not possible in linear systems.

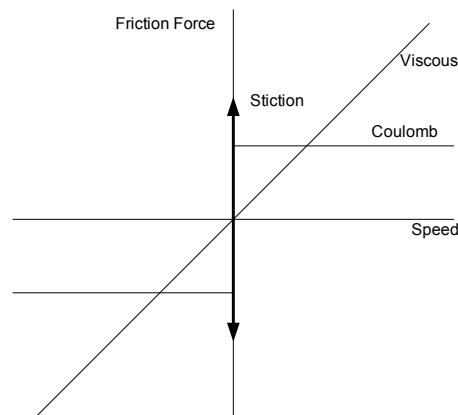
## 1.2 Forms of nonlinearity

All practical systems are nonlinear and in this section a brief overview is given of some nonlinear effects that often occur. In initial designs it may be possible to approximate the nonlinear effects by linear models but invariably it will be necessary to finally check their effects on the system performance either in simulation or/and the real hardware. Today's digital simulation languages are very good but to use them efficiently for investigating the effects of nonlinearity, or to assist in the design of a nonlinear system, requires a knowledge of the supporting theoretical methods presented in this book.

In typical control engineering problems nonlinearity may occur in the dynamics of the plant to be controlled or in the components used to implement the control. In the latter case, for example, a valve actuator may have a dead zone due to friction effects and will certainly saturate for large inputs, so this may be referred to as an inherent nonlinearity, because it exists although one might possibly prefer this not to be the case. Alternatively one may have intentional nonlinearities which have been purposely designed into the system to improve the system specifications, either for technical or economic reasons. A good example of this is the on-off control used in many temperature control systems, where the objective is to have the temperature oscillate about the required value.

Identifying the precise form of a nonlinearity 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. It might be argued that linear systems theory is not applicable to practical control engineering problems because they are always nonlinear. This is an overstatement, of course, but a valid reminder. All systems have actuator saturation and in some cases it might occur for relatively low error signals, for example in rotary position control it is not unusual for a step input of say  $10^\circ$ , or even less, to produce the maximum motor drive torque. It simply is the result of good economical design as to produce linear operation to a higher torque level would require a larger and more expensive motor. Valves used to control fluid or gas flow, apart from having the nonlinear effects mentioned above, can have a slightly different behavior when opening compared with closing due to the unidirectional pressure of the fluid.

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, illustrated in Figure 1.2, 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 shaft loads, for example a fan for which the drive torque typically increases as a power of speed.



**Figure 1.2** Three basic friction components

A mathematical expression sometimes used to approximate friction is

$$n(v) = (a - be^{-c|v|} + d|v|)\operatorname{sgn}(v) \quad (1.3)$$

Here the parameters  $a$ ,  $b$ ,  $c$  and  $d$  are chosen to provide the desired shape against speed which initially decreases from the value at zero speed and then starts to increase.

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, which are only available for low torques. Backlash, which is a very complicated phenomenon, involving impacts between surfaces is often modeled in a very simplistic manner. For example, the simple approach used in some digital simulation languages, such as Simulink, the simulation component of the well-known software package, Matlab, 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 but no checks exist in Simulink to determine how good they are or, indeed, when they are completely invalid. Better facilities for modeling phenomena such as backlash can be found in simulation languages such as 20-Sim or Dynast, which do not use the block concept of Simulink. A good discussion of friction and backlash can be found in reference 1.1.


Probably, the most widely used intentional nonlinearity is the relay. The on-off type, which can be described mathematically by the signum function, that is switches on if its input exceeds zero and off if it goes below zero, 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 the temperature control of buildings, where typically the switching is provided from a temperature sensor 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 and closes a contact. When the temperature increases the coil expands causing a change in angle for the mercury to flow and break the contact.

Electronic switching controllers are being 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. When used in a position control system the zero output allows for a steady state position within the dead zone but this affects the resulting steady state control accuracy.


Hysteresis effects in magnetic materials sometimes have to be modeled. This is often done by assuming a hysteresis loop of the form of a B-H loop for a magnetic material typically obtained for a sinusoidal input. However the shape usually varies with the amplitude and frequency of the input and does not in fact remain constant with a random excitation. Provided the input is sinusoidal and the shape does remain reasonably constant then a nonlinear function of the form  $n(x, \dot{x})$  may provide a reasonable model as the path taken around the hysteresis depends upon whether the input,  $x$ , is increasing or decreasing, i.e the derivative of  $x$ ,  $\dot{x}$ .

Please click the advert

SIMPLY CLEVER



## We will turn your CV into an opportunity of a lifetime



Do you like cars? Would you like to be a part of a successful brand?  
We will appreciate and reward both your enthusiasm and talent.  
Send us your CV. You will be surprised where it can take you.

Send us your CV on  
[www.employerforlife.com](http://www.employerforlife.com)





### 1.3 Structure and Behaviour

From a control engineering viewpoint there are two major reasons why one needs to know about nonlinearity. Firstly with respect to obtaining mathematical models of devices, particularly if identification techniques are being used, and secondly for ensuring the design meets the desired specifications when the control system is nonlinear. To appreciate these aspects it is appropriate to discuss very briefly in the introduction a few aspects of behaviour due to the presence of nonlinearity. These are dependent on the structure of the nonlinear system and the relevance of this is explained by again considering a sinusoidal input.

Consider a nonlinearity  $x + dx^3$ , and a dynamic element with transfer function  $K/s(s + b)$ , where  $b$  and  $d$  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 any given discrete input frequency spectrum, which in principle can be evaluated. New frequency components can only be created by the nonlinearities and the linear elements simply alter the relative magnitudes and phases of these components.

For example, if the above nonlinearity is placed before and after the 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 magnitudes and phases dependent on both the input sinusoidal magnitude and 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 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 input sinusoid. Thus an approximate frequency response 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 of the linear dynamic elements, like the one given, have low pass dynamics so that the frequency,  $f$ , will predominate 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 the input amplitude. This representation of a nonlinear element is known as a describing function (DF), which is the topic of several chapters. Sometimes the above model for the combination based on the fundamental frequency,  $f$ , only, is referred to as an amplitude and frequency dependent describing function.

If alternatively we assume the system structure to consist of a feedback combination of the linear and nonlinear elements, with the transfer function  $K/s(s + b)$  in the forward loop and the feedback loop containing  $x + dx^3$  fed back through a negative gain then, a very different situation is possible for the response to a sinusoidal input. Dependent on the values of  $b$ ,  $d$  and the amplitude and frequency of the input, some possibilities for the output are that it is (a) approximately sinusoidal with the same frequency as the input, similar to the aforementioned cascade connection; (b) approximately sinusoidal with a frequency related to that of the input; (c) a combination of primarily the input frequency and another frequency or (d) a waveform known as chaotic, which is not definable mathematically but completely repeatable for the same initial conditions.

These behaviours and others are unique to nonlinear feedback systems, aspects which make such systems extremely interesting. However, it has meant that no general analytical method is available for predicting their behavior. Several approaches will be considered in this book all of which will be restricted in their applicability or, put alternatively, the situations which they can address. Thus the importance of simulation studies for investigating nonlinear systems in association with analytical methods cannot be underestimated. Much of the support for the theoretical material presented in the early chapters, particularly in the 50s to 60s, was done using analogue simulation. Today simulations are done digitally and several are included using Simulink in the following chapters to illustrate the concepts and provide solutions for specific problems. Care has to be taken in simulating nonlinear systems particularly those with linear segmented characteristics because of the discontinuities. Some comments are made on the simulations where appropriate.

## 1.4 Overview of contents

The phase plane approach discussed in chapter 2 is very useful for step response and stability studies but is basically restricted to second order systems. However, many engineering systems, particularly in the mechatronics field, may be approximated by a second order differential equation so the results are still of value. It also provides a simple basis for understanding some of the more advanced topics, such as optimum control and sliding mode control, covered in chapter 9.

Stability of a feedback loop is of major importance so that it is not surprising that much early work was concerned with this topic. During the 1940's engineers in several countries developed what has become known as the describing function method where a nonlinearity is replaced by an amplitude dependent gain to a sinusoid, known as the describing function, DF. Chapter 3 introduces the DF, shows how its value can be calculated for various nonlinearities and includes a table of results. The next two chapters, 4 and 5, deal with applications of the describing function for estimating limit cycles and the stability of a nonlinear feedback loop. It is also shown how the describing function can be evaluated for other than a single sinusoidal input and applications of describing functions for a bias plus sinusoid and two sinusoids are given. These include areas such as limit cycle stability, jump resonance and subharmonic oscillations.

The relay is a special form of nonlinear element where the output is not continuously dependent on the input. This allows special results to be developed for relay systems a topic covered in chapter 6. Chapter 7 deals with a design approach that has received much attention in recent years, namely using the information contained in relay produced limit cycles to set the parameters of a controller, typically known as relay autotuning. The topic of absolute stability, namely the development of exact criteria for guaranteeing the stability of a feedback loop with a single nonlinearity and linear transfer function, is covered in chapter 8. Dependent on the viewpoint these results may be regarded as robust, as they prove stability for variations in the nonlinearity, or conservative if one is considering stability for a specific nonlinearity.

The coverage to this point, apart from chapter 7, has like many textbooks on linear control, been primarily concerned with introducing analytical tools. Chapter 9, however, focuses more on design and looks at how some of the ideas covered and other techniques may be used for nonlinear control system design. The coverage of these topics is quite brief, some necessarily so because of the additional theoretical concepts which would have to be introduced to go into them more deeply. Hopefully, sufficient information, together with the references, is given for the reader to understand the concepts involved, their possible relevance for particular applications and how they might be applied.

## 1.5 References

1.1 Friedland, B, 1996 Advanced control system design, Chapter 7 Prentice Hall.

Please click the advert

With us you can  
shape the future.  
Every single day.

For more information go to:  
[www.eon-career.com](http://www.eon-career.com)

Your energy shapes the future.

**e-on**

## 2 The Phase Plane Method

### 2.1 Introduction

A significant amount of the research into nonlinear differential equations in the nineteenth and early twentieth centuries done by mathematicians and physicists was devoted to second order differential equations. There were two major reasons for this, namely that the dynamics of many problems of practical interest could be approximated by these equations and secondly the phase plane approach allowed a graphical examination of their solutions. The systems of interest in the late nineteenth and early twentieth centuries were found in fields such as celestial mechanics, nonlinear mechanical systems and electronic oscillations. This section will introduce the basic concepts of the phase plane approach and then give a brief overview of how the method has been further developed for use in control system analysis and design.

A significant amount of the early development in control theory from 1930 was driven by two areas, namely to achieve better control of industrial processes and to achieve better performance in fire control problems. The problems in the latter area received significantly more attention in the war years after 1939, where the requirement in many cases was related to the position control of radar antennas and guns in both stationary and moving situations. The dynamic equations representing many of these position control systems could be represented reasonably accurately by second order nonlinear differential equations. It was therefore not surprising that much of the early work on nonlinear control used the phase plane approach. Control engineers did make significant contributions to this field since, whereas the earlier work had typically assumed nonlinearities defined by continuous mathematical functions, for control system analysis it was often more appropriate to approximate intrinsic nonlinearity, such as friction, or intentionally introduced nonlinearity, such as a relay, by linear segmented characteristics. The approach is still useful today because of the fundamental insight it provides into aspects of nonlinear system behaviour; the fact there are still many control problems which can be approximated by second order dynamics, and also because more than one nonlinearity can be considered.

### 2.2 Basic Principles

The formulation used in early work on second order systems was to assume a representation in terms of the two first order equations

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

Equilibrium, or singular, points represent a stationary system for the dynamics and occur when  $\dot{x}_1 = \dot{x}_2 = 0$

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

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

Before considering the nonlinear case further it is important to first examine the linear case, which is rarely discussed in detail in texts on linear control. Most, usually, just restrict the contribution to considering the step response of a second order system.

### 2.2.1 The Linear Case

The equations for the linear situation may be written

$$\begin{aligned}\dot{x}_1 &= ax_1 + bx_2 \\ \dot{x}_2 &= cx_1 + dx_2\end{aligned}\tag{2.3}$$

or in matrix form

$$\begin{aligned}\dot{x} &= Ax \\ \text{where } A &= \begin{pmatrix} a & b \\ c & d \end{pmatrix}.\end{aligned}\tag{2.4}$$

Eliminating  $x_2$  from equations (2.3) by differentiating the first equation and substituting for  $\dot{x}_2$  from the second equation and  $x_2$  from the first yields

$$\ddot{x}_1 - (a + d)\dot{x}_1 + (ad - bc)x_1 = 0$$

which after taking Laplace transforms has the characteristic equation

$$s^2 - (a + d)s + ad - bc = 0$$

which is the same as the eigenvalue equation of the matrix  $A$ , with  $\lambda$  typically replacing  $s$  when discussed in mathematical texts.

The roots, or eigenvalues, of the equation depend on the values of the elements of  $A$ . If the equation is written in the form

$$\lambda^2 + B\lambda + C = 0$$

then the two eigenvalues are given by

$$\lambda_1 = -B + (B^2 - 4C)^{1/2}/2 \text{ and } \lambda_2 = -B - (B^2 - 4C)^{1/2}/2$$

The solution to the differential equation is of the form

$$x_1(t) = K_1 e^{\lambda_1 t} + K_2 e^{\lambda_2 t} \text{ for } \lambda_1 \neq \lambda_2$$

$$x_1(t) = K_1 e^{\lambda_1 t} + tK_2 e^{\lambda_1 t} \text{ for } \lambda_1 = \lambda_2$$

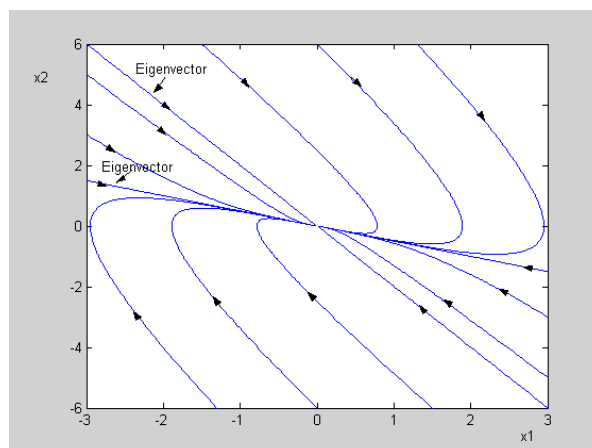
For the linear system there is only one singular point at the origin  $x_1 = 0, x_2 = 0$  of the state plane and the behaviour of the motion near to the singular point depends on the eigenvalues.

Trajectories cannot intersect and can only meet at a singular point. The following four cases can occur for the singular point.

- (i)  $\lambda_1$  and  $\lambda_2$  are both real and have the same sign (either positive or negative).
- (ii)  $\lambda_1$  and  $\lambda_2$  are both real and have opposite signs.
- (iii)  $\lambda_1$  and  $\lambda_2$  are complex conjugates with non zero real parts.
- (iv)  $\lambda_1$  and  $\lambda_2$  are imaginary, that is complex conjugates with zero real parts.

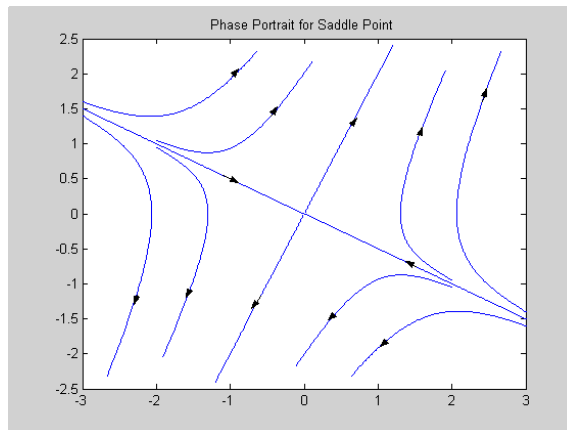
Case (i) corresponds to a singular point known as a node, which is stable when the eigenvalues are negative and unstable when they are positive. The form of the trajectories near to the singular point is shown in the simulation results of Figure 2.1, for the transfer function  $1/(s^2 + 2.5s + 1)$  which has eigenvalues of -0.5 and -2, with corresponding eigenvectors of  $(1, -2)^T$  and  $(1, -0.5)^T$ .

Arrows are typically placed on the trajectories showing the direction of motion which will be towards the singular point for a stable node, as shown in the figure, and away from it for an unstable one. Trajectories starting on an eigenvector remain on it as is clearly seen in the figure. For the stable node all trajectories, apart from that starting on the first eigenvector tend towards the node along the second eigenvector, i.e the one with the smallest slope. The plot showing trajectory motions from different initial conditions is known as a state or phase portrait, although the latter name is often reserved for the special case where  $\dot{x}_1 = x_2$ , that is  $a = 0$  and  $b = 1$  in equation (2.3)



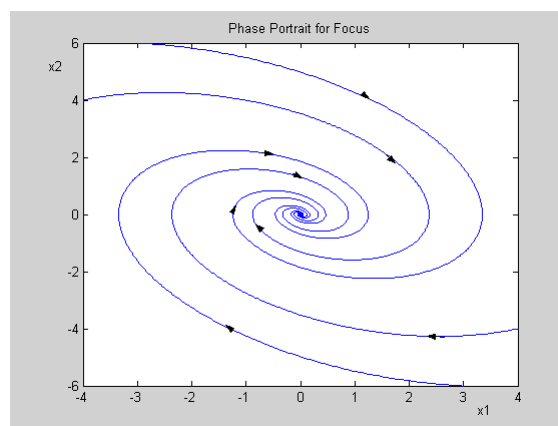
**Figure 2.1** Phase Portrait for a Node

The singular point in case (ii) is a saddle point which is only theoretically reachable from initial conditions on the eigenvector of the stable solution associated with the negative eigenvalue. Any small perturbation will cause the trajectories to diverge as shown by the phase portrait of Figure 2.2 and move outwards as shown by the arrows.



**Figure 2.2** Phase Portrait for a Saddle Point

A phase portrait for case (iii) is shown in Figure 2.3 where the singular point is a focus. The focus is stable when the real part is negative, so that trajectories spiral towards it as marked in the figure, and unstable when the real part is positive. Four trajectories are shown in the figure.



**Figure 2.3** Phase Portrait for a Focus Showing Four Trajectories

In the final case (iv) the singular point is a centre, with the oscillatory motion producing concentric trajectories, as shown in Figure 2.4. Their size depends on the initial conditions, which in a physical situation, such as an ideal oscillating pendulum, is defined by the initial input energy.



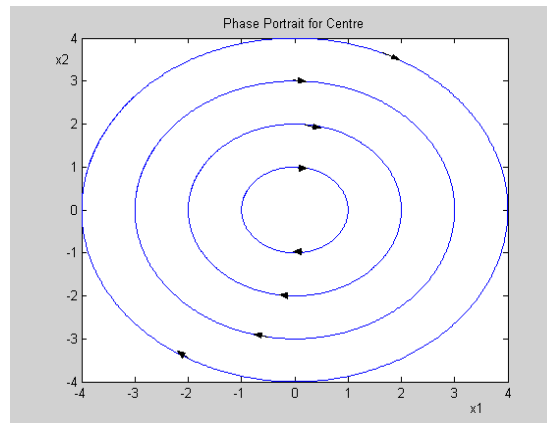


Figure 2.4 Phase Portrait for a Centre

### 2.2.2 The Nonlinear Case

In general it may not be possible to obtain analytical solutions to even second order nonlinear differential equations so the value of the phase plane approach in early work was to allow approximate solutions to be obtained using graphical techniques for sketching phase plane trajectories. The basic approach used was to determine the slope at a sufficient number of points in the state plane to allow a picture of the motion to be obtained starting from any initial conditions in the  $x_1 - x_2$  state plane. From equation (2.2) a trajectory will have a slope  $r$  when it crosses the curve

$$rP(x_1, x_2) + Q(x_1, x_2) = 0 \quad (2.5)$$

Please click the advert

**Nido**

**Luxurious accommodation**

**Central zone 1 & 2 locations**

**Meet hundreds of international students**

**BOOK NOW and get a £100 voucher from voucherexpress**

**Nido Student Living - London**

**Visit [www.NidoStudentLiving.com/Bookboon](http://www.NidoStudentLiving.com/Bookboon) for more info.**

**+44 (0)20 3102 1060**

By selecting a range of values of  $r$ , drawing the corresponding curves and marking the slope  $r$  with a short arrow at which a trajectory crosses, allows phase portraits to be sketched. This is usually known as the method of isoclines. Sometimes the lengths of the arrows are drawn in proportion to the velocity at the point and the plot is then called a vector field plot.

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

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

and if this is rearranged as two first order equations, choosing the phase variables, that is the output and its derivative, as the state variables, then  $x_1 = x$ ,  $x_2 = \dot{x}$  and the equation can be written as

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= -f(x_1, x_2)\end{aligned}\tag{2.6}$$

which is a special case of equation (2.2). It may have multiple singular points but they will all lie on the  $x_1$  axis, that is for  $x_2 = 0$ .

A phase portrait can therefore be obtained for any function,  $f$ , using the method of isoclines, or possibly other methods which can be found in the literature [2.1]. Digital simulation methods now enable phase plane trajectories to be easily displayed and they can often provide a clearer picture of the system behaviour than time response plots for  $x_1$  and  $x_2$ . Many situations in science and engineering may be modeled approximately by equation (2.1) or the special case equation (2.6). Some good examples can be found in the recent book given in reference [2.2] whilst here just one well known example is taken.

### 2.2.3 An Example

Many investigations using the phase plane technique were concerned with the possibility of a nonlinear differential equation having a limit cycle solution, the strict name for an oscillation in a nonlinear system. When a limit cycle exists it results in a closed trajectory in the phase plane and typical of such investigations was the work of Van der Pol on oscillators. He considered the equation

$$\ddot{x} - \mu(1 - x^2)\dot{x} + x = 0\tag{2.7}$$

where  $\mu$  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

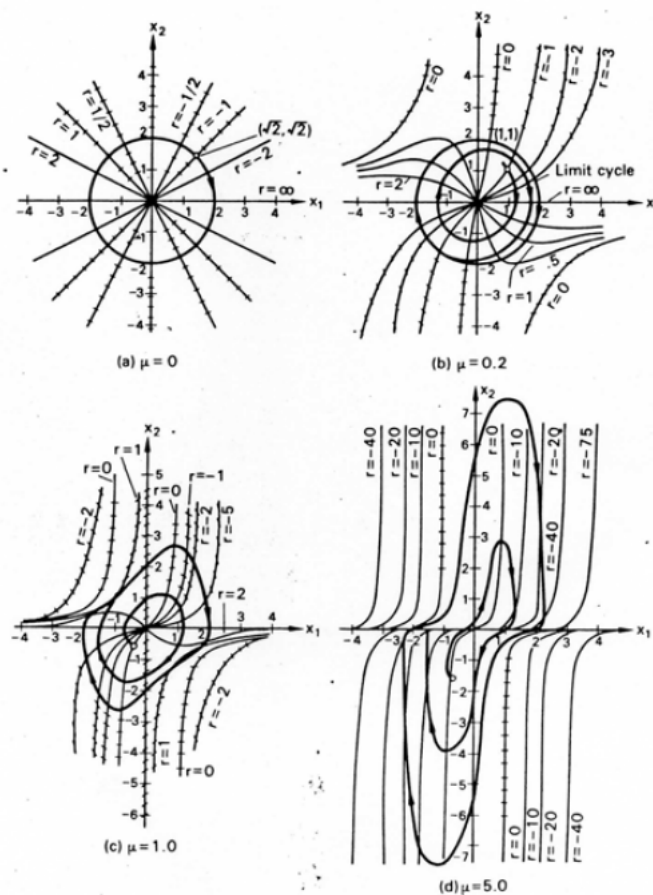
$$\frac{dx_2}{dx_1} = \frac{\dot{x}_2}{\dot{x}_1} = \frac{\mu(1 - x_1^2)x_2 - x_1}{x_2}\tag{2.8}$$

It has one singular point at the origin  $(0, 0)$  about which the linearised equation is  $\ddot{x} - \mu\dot{x} + x = 0$ , so that the singular point depends upon the value of  $\mu$ . It is an unstable focus for  $\mu < 2$  and an unstable node for  $\mu > 2$ .

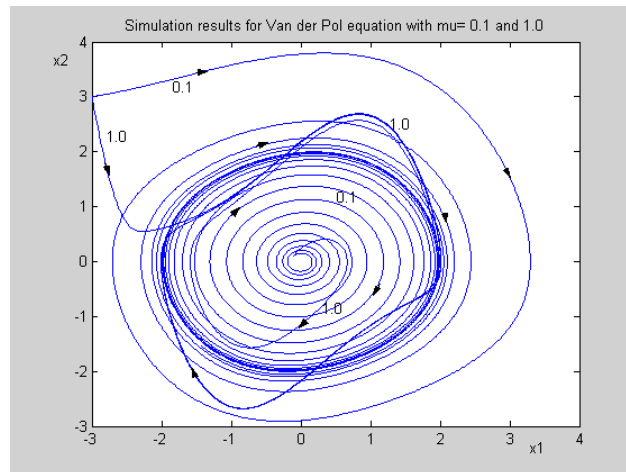
All phase plane trajectories have a slope of  $r$  when they intersect the curve

$$rx_2 = \mu(1 - x_1^2)x_2 - x_1 \quad (2.9)$$

Figure 2.5 shows phase portraits obtained using isocline sketching and Figure 2.6 shows simulation results obtained from the initial conditions of  $(-3, 3)$  and  $(-0.1, 0.1)$  for  $\mu = 0.1$  and  $1$ . It can be seen, as expected, that the resulting limit cycles are dependent on  $\mu$  with that for the small value of  $\mu$  being almost elliptical. They are reached from outside for the first initial condition and from inside for the second initial condition. A limit cycle has this feature that it is reached from different initial conditions and in this case it is stable as trajectories tend to it from both without and within. In contrast if  $\mu = 0$  then one has an oscillation, not a limit cycle, and the magnitude of the oscillation is determined by the initial conditions. Since the system has no damping there is no loss of energy and its value is set by the initial conditions.



**Figure 2.5** Phase portraits of the Van der Pol equation from isocline sketching



**Figure 2.6** Phase plane plots of the Van der Pol equation for two values of

### 2.3 The Phase Plane for Systems with Linear Segmented Nonlinearities

It is an advantage when using the phase plane approach to have nonlinearities which are described by linear segmented characteristics. This is because it results in a phase plane which can be divided up into different regions with different linear differential equations describing the motion in each region. Although the differential equations are linear the mathematical phase plane equations are only simple enough for calculating analytical solutions by 'hand' in a few cases. To illustrate these basic concepts three examples are considered in the next section. Simple dynamics are used in some cases to allow analytical solutions for the motion but also shown are some simulations with other dynamics.

Please click the advert

I joined MITAS because  
I wanted **real responsibility**

The Graduate Programme  
for Engineers and Geoscientists  
**Maersk.com/Mitas**



**Month 16**

I was a construction  
supervisor in  
the North Sea  
advising and  
helping foremen  
solve problems

Real work  
International opportunities  
Three work placements



**MAERSK**

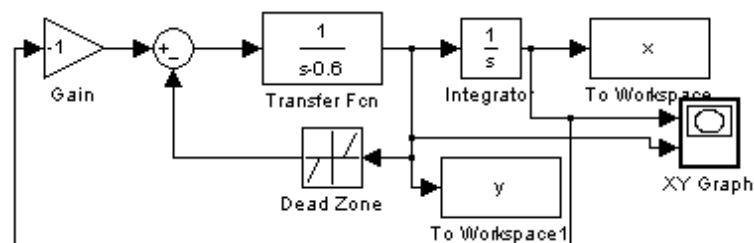
Download free ebooks at [bookboon.com](http://bookboon.com)

### 2.3.1 Example 1 – Nonlinear Output Derivative Feedback

The Simulink diagram for this example is given in Figure 2.7, where the only nonlinearity is in the derivative of output feedback path. To plot the phase plane using Simulink the X-Y oscilloscope can be used or alternatively the output and its derivative can be fed to simout blocks for plotting after the simulation run has finished. The simout blocks must be put in the array mode and the time vector is available in tout. Often using the default integration algorithm in Simulink waveforms do not appear smooth so one may have to use an algorithm with a small fixed step size or limit the step size in an algorithm. Results from the X-Y scope can be read to the display accuracy. More accurate results can be obtained from the simout blocks but plotting the data may not be straightforward as one may have unequal vector lengths of data stored when a default integration algorithm is not used.

If the nonlinearity is replaced by a gain  $K$  then the characteristic equation is

$$s^2 + (K - 0.6)s + 1 = 0 \quad (2.10)$$

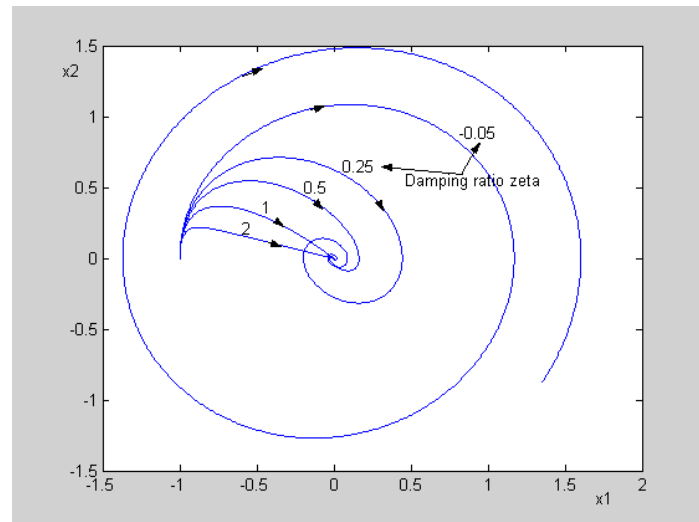


**Figure 2.7** Simulink diagram for system with dead zone feedback

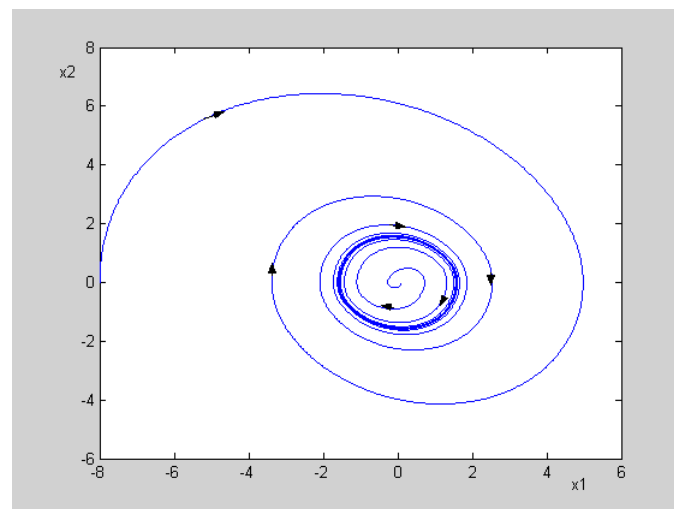
It is clear that for  $K < 0.6$  the linear system is unstable and stable for  $K > 0.6$ . For the dead zone the gain is zero for a small sinusoidal signal but approaches unity for a large sinusoidal signal. Thus, one expects the effect of the dead zone nonlinearity to change the damping in some way such that the system is unstable for small signals and stable for large signals. This suggests the existence of a stable limit cycle. Figure 2.8 shows phase plane plots starting from  $(-1, 0)$  for the linear second order system with characteristic equation

$$s^2 + 2\zeta s + 1 = 0 \quad (2.11)$$

for various values of  $\zeta$ , the damping ratio. For the unstable case with  $\zeta = -0.05$  the trajectory slowly spirals outwards and for the stable responses they converge to the origin in a spiral manner for  $\zeta < 1$ , when the origin is a focus and directly with no overshoot for  $\zeta \geq 1$ , when the origin is a node. The mathematical expressions for these curves are quite complicated, so to sketch them on a phase plane the method of isoclines or some special methods may be used [2.1]. The dead zone is set at levels of  $\pm 0.5$ , so that the trajectories describing the motion are those of a linear second order system with a damping ratio of  $-0.3$  for  $|x_2| < 0.5$  and  $0.2$  for  $|x_2| > 0.5$ . Phase plane plots have been obtained by simulation and are shown from initial conditions of  $(-8, 0)$  and  $(0.1, 0)$  in Figure 2.9. The resulting limit cycle is clearly shown in the figure.



**Figure 2.8** Phase plane plots for second order system

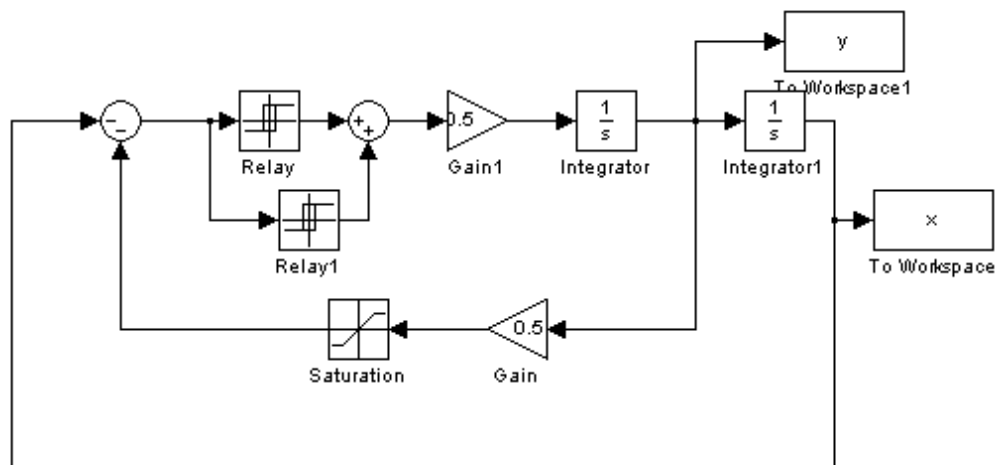


**Figure 2.9** Phase plane plots for system of Figure 2.7

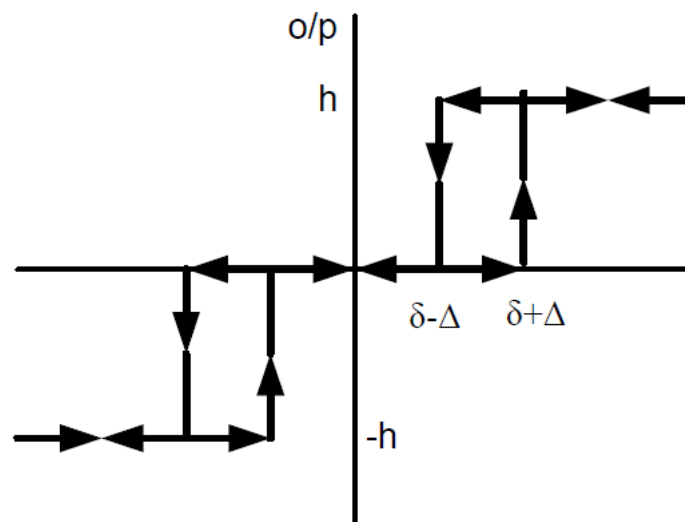
### 2.3.2 Example 2 – Relay Position Control

Consider a basic relay position control system with nonlinear velocity feedback having the Simulink diagram shown in Figure 2.10. The two relays in parallel simulate a relay with dead zone and hysteresis which has the characteristic shown in Figure 2.11, where for positive inputs the switch on level is  $\delta + \Delta$  and the switch off level is  $\delta - \Delta$ . The relay parameters are set to give an output,  $h$ , of  $\pm 1$ , and 0.





**Figure 2.10** Simulink diagram of the relay control system for example 2



**Figure 2.11** Relay with dead zone and hysteresis

It is assumed initially that the hysteresis in the relay is negligible (i.e.  $\Delta = 0$ ) and that there is no saturation in the velocity feedback path.. Denoting the system position output by  $x_1$  and its derivative  $\dot{x}_1$  by  $x_2$  then the feedback to the relay input is assumed to be  $-x_1 - \lambda x_2$ . The relay output of  $\pm 1$  or  $0$  is equal to  $\ddot{x}_1/K$ , where in the Simulink diagram both  $K$  and  $\lambda$  are chosen equal to  $0.5$ . Taking the dead zone of the relay  $\pm\delta$  to be equal to  $\pm 1$ , the motion of the system is described by

$$\ddot{x}_1 = \begin{cases} K & \text{if } -x_1 - \lambda x_2 > 1 \\ 0 & \text{if } |-x_1 - \lambda x_2| < 1 \\ -K & \text{if } -x_1 - \lambda x_2 < -1 \end{cases} \quad (2.12)$$

Thus the equation of motion in the phase plane changes at the lines  $x_1 + \lambda x_2 = \pm 1$  and these can be drawn on the plot, as shown in Figure 2.12, to divide up the phase plane into the three regions where the motion is described by the above three simple linear second-order differential equations.



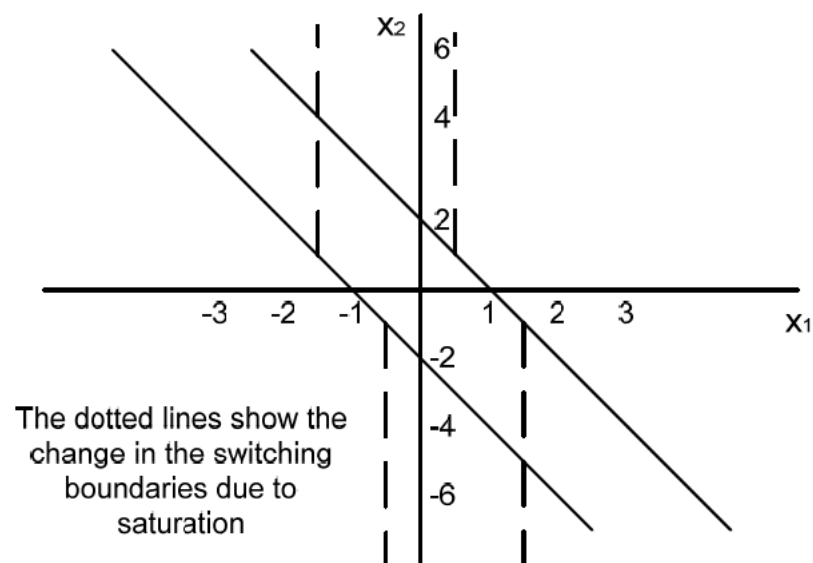


Figure 2.12 Switching lines in the phase plane

The solution of

$$\ddot{x}_1 = K \quad (2.13)$$

Please click the advert

# Brain power

Plug into The Power of Knowledge Engineering.  
Visit us at [www.skf.com/knowledge](http://www.skf.com/knowledge)

By 2020, wind could provide one-tenth of our planet's electricity needs. Already today, SKF's innovative know-how is crucial to running a large proportion of the world's wind turbines.

Up to 25 % of the generating costs relate to maintenance. These can be reduced dramatically thanks to our systems for on-line condition monitoring and automatic lubrication. We help make it more economical to create cleaner, cheaper energy out of thin air.

By sharing our experience, expertise, and creativity, industries can boost performance beyond expectations. Therefore we need the best employees who can meet this challenge!

**The Power of Knowledge Engineering**

is obtained by writing it in the form

$$x_2 \frac{dx_2}{dx_1} = K \quad (2.14)$$

and integrating with the initial conditions for  $x_1$  and  $x_2$  of  $x_{10}$  and  $x_{20}$ , respectively, to give

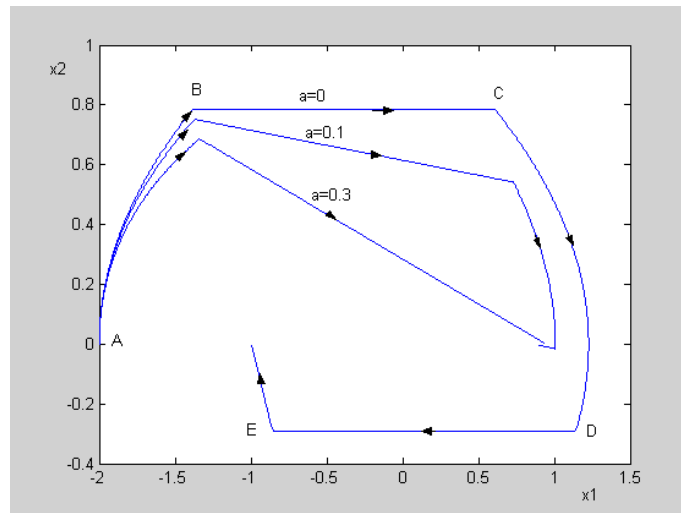
$$x_2^2 - x_{20}^2 = 2K(x_1 - x_{10}) \quad (2.15)$$

where  $x_{10}$  and  $x_{20}$  are the initial values of  $x_1$  and  $x_2$ . Since equation (2.15) describes a parabola, which for the special case of  $K = 0$  has the solution  $x_2 = x_{20}$ , it is easy to calculate the system's response from any initial condition  $(x_{10}, x_{20})$  in the phase plane.

Figure 2.13 shows the response from  $(-2, 0)$  with  $\lambda = K = 0.5$  as given in Figure 2.7. The initial parabola, with  $K = 0.5$ , starts from A and meets the first switching boundary  $x_1 + 0.5x_2 = -1$  at B; the ensuing motion is horizontal, that is, at constant velocity, until the second switching boundary  $x_1 + 0.5x_2 = 1$  is reached at C. The ensuing parabola, with  $K = -0.5$ , meets the same switching boundary again at D. The next motion is again at constant velocity until the point E is reached on the first switching boundary. At this point the following parabolic motion, with  $K = 0.5$ , brings the trajectory straight back to the switching boundary. The situation has therefore arisen where the motion at either side of the switching line is directed back to it. Thus, the resulting motion, which is known as a sliding mode, is directed along the switching line to come to rest at  $(-1, 0)$ . In theory switching takes place at an infinite rate but in practice the relay will have some hysteresis, which will decrease the switching rate.

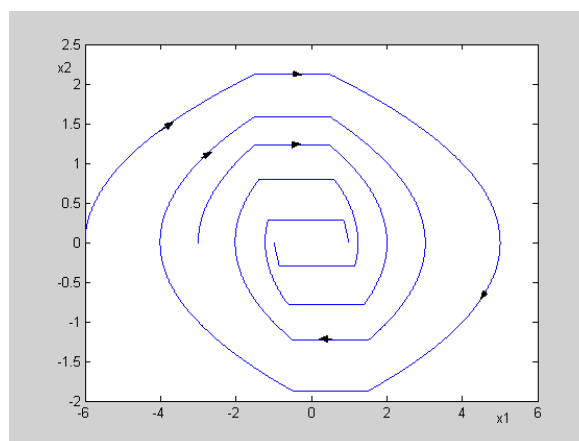
By solving the relevant equations it can be shown that  $B = (-1.39, 0.781)$ ,  $C = (0.610, 0.781)$ ,  $D = (1.14, -0.281)$  and  $E = (-0.86, -0.281)$ . If the first integrator is replaced by the transfer function  $1/(s + a)$  to provide some damping then the analytical solution is not as easy. The motion in the phase plane when the relay output is zero is still linear, but now with a slope of  $-1/a$ , and when the relay output is  $\pm 1$  it is no longer parabolic. Simulation responses are also shown in Figure 2.13 for values of  $a$  equal to 0.1 and 0.3. Responses from any other initial conditions are obviously easy to find, but, from the responses 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  $\pm 1$ . Thus a steady-state error of unit magnitude will result from any motion.

Finally a word of warning that difficulties may arise in simulating systems with sliding, since if a variable step length integration algorithm is used it may reduce the step length to zero and stop in the sliding motion. Many simulation languages have special routines for taking care of the discontinuous sliding motion, but good results can usually be obtained by using a fixed step length integration algorithm or by incorporating a small hysteresis in the relay.



**Figure 2.13** Initial condition response for system with

When the velocity feedback signal saturates, that is, when  $|\lambda x_2| > h$ , the input signal to the relay is  $-x_1 \pm h$ . Thus, if  $h$  is chosen equal to 0.5 when  $|x_2| > 1$  the switching boundaries become vertical lines as shown dotted in Figure 2.12, although the equations describing the motion between the boundaries remain unaltered. Therefore for a large step input of  $r$ , the equivalent of starting from the initial condition  $(-r, 0)$ , the response will become more oscillatory when the velocity saturates as illustrated in Figure 2.14 for values of  $r$  equal to 6 and 3 and saturation of the velocity feedback signal at  $\pm 0.5$  for inputs greater than  $\pm 0.5$ . The parabolic curves are clearly seen, since the damping has been taken as zero, as also is the fact that the switching lines have become vertical for  $|x_2| > 1$ .



**Figure 2.14** Phase plane plots for step inputs of 6 and 3 with velocity saturation but no hysteresis or damping.

Again if there is no velocity saturation in the feedback path but the relay has a finite hysteresis,  $\Delta$ , then the switching boundaries are again straight lines of slope  $-1/\lambda$  but different according to whether  $x_2$  is positive or negative. In the general case they are

$$\begin{aligned} x_1 + \lambda x_2 &= -\delta + \Delta \quad \text{and} \quad x_1 + \lambda x_2 = \delta + \Delta \quad \text{for } x_2 \text{ positive, and} \\ x_1 + \lambda x_2 &= -\delta - \Delta \quad \text{and} \quad x_1 + \lambda x_2 = \delta - \Delta \quad \text{for } x_2 \text{ negative.} \end{aligned} \quad (2.16)$$

It is easily shown from symmetry considerations, when the above lines are drawn on a phase plane, that a limit cycle exists for the system with no damping with a maximum value of  $x_2 = \Delta/\lambda$ . Figure 2.15 shows simulation results for  $\delta = 1$  and  $\lambda = \Delta = 0.5$ . The responses from -1.51 and -4 are shown converging to the limit cycle from within and without. If the system were stabilized by replacing the first integrator by the transfer function  $1/(s + a)$  then the motion could come to rest, dependent on the input step magnitude, with the relay input in the range  $\pm 1.5$ .

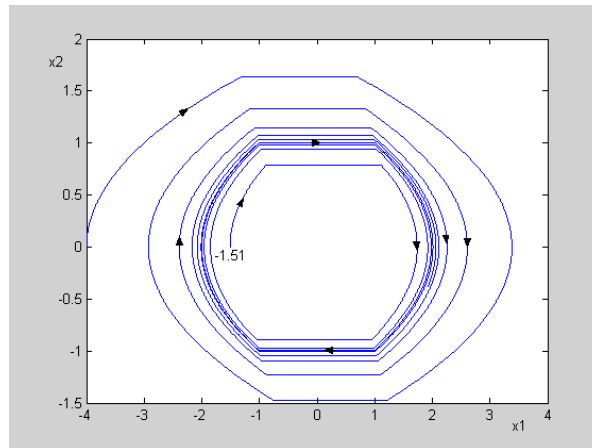


Figure 2.15 Initial condition responses resulting in a limit cycle

Please click the advert

## Are you considering a European business degree?

**LEARN BUSINESS** at university level. We mix cases with cutting edge research working individually or in teams and everyone speaks English. Bring back valuable knowledge and experience to boost your career.

**MEET** a culture of new foods, music and traditions and a new way of studying business in a safe, clean environment – in the middle of Copenhagen, Denmark.

**ENGAGE** in extra-curricular activities such as case competitions, sports, etc. – make new friends among CBS' 18,000 students from more than 80 countries.



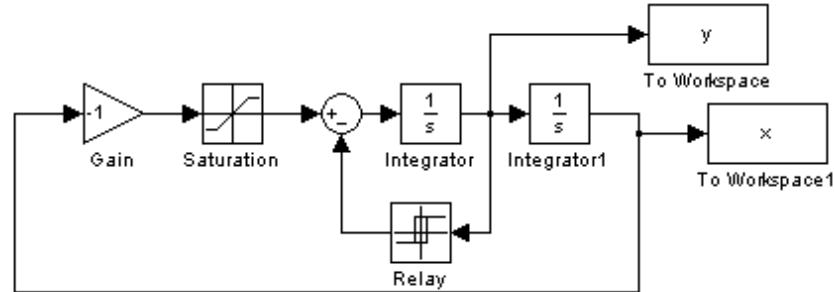
See what we look like  
and how we work on [cbs.dk](http://cbs.dk)



Download free ebooks at [bookboon.com](http://bookboon.com)

### 2.3.3 Example 3 – Position Control with Torque Saturation

Figure 2.16 shows a Simulink diagram for a position control system with no viscous damping and with nonlinear effects due to torque saturation and Coulomb friction.



**Figure 2.16** Simulink diagram of control system for Example 3

The differential equation of motion in phase variable form has

$$\dot{x}_2 = f_s(-x_1) - 0.5\text{sgn}(x_2) \quad (2.17)$$

where  $f_s$  denotes the saturation nonlinearity and  $\text{sgn}$  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, equation (2.17) can be written

$$\ddot{x}_1 + f_s(x_1) + 1/2 = 0 \quad (2.18)$$

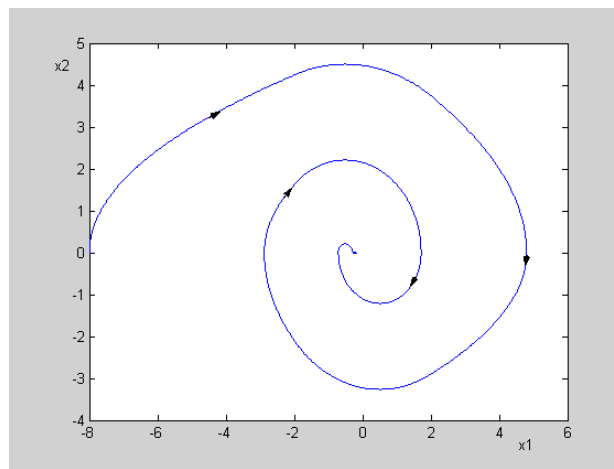
so that for

- (a)  $x_2 + ve, x_1 < -2$ , one has  $\dot{x}_1 = x_2, \dot{x}_2 = 3/2$ , a parabola in the phase plane.
- (b)  $x_2 + ve, -2 \leq x_1 \leq 2$ , one has  $\dot{x}_1 = x_2, \dot{x}_2 + x_1 + 1/2 = 0$ , a circle in the phase plane.
- (c)  $x_2 + ve, x_1 > 2$ , one has  $\dot{x}_1 = x_2, \dot{x}_2 = -5/2$ , a parabola in the phase plane.

Similarly for  $x_2$  negative,

- (d)  $x_2 - ve, x_1 < -2$ , one has  $\dot{x}_1 = x_2, \dot{x}_2 = 5/2$ , a parabola in the phase plane.
- (e)  $x_2 - ve, -2 \leq x_1 \leq 2$ , one has  $\dot{x}_1 = x_2, \dot{x}_2 + x_1 - 1/2 = 0$ , a circle in the phase plane.
- (f)  $x_2 - ve, x_1 > 2$ , one has  $\dot{x}_1 = x_2, \dot{x}_2 = -3/2$ , a parabola in the phase plane.

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



**Figure 2.17** Response from initial condition  $(-6, 0)$  for Example 3

Careful examination of Figure 2.17, which shows the simulation result for a response from  $(-6, 0)$ , reveals the changes in the trajectory shape when crossing the lines  $x_1 = \pm 2$ .

## 2.4 Conclusions

This chapter has covered the basic concepts in the analysis of second order systems using the phase plane approach. Since many simple control systems can be approximated by nonlinear second order differential equations the approach is powerful. It can be used when more than one nonlinear element exists and is particularly useful when the nonlinearity can be approximated by linear segmented characteristics. Examples have been given using a double integrator plant transfer function with a constant input, which results in simple mathematical expressions for the phase plane trajectories, so that certain features can be easily illustrated.

Mathematical expressions for the trajectories can always be obtained, since they are given by the solutions of linear differential equations when the nonlinearities are linear segmented characteristics, but they are often quite complicated. Phase plane plots can easily be obtained with modern simulation languages but the facility to sketch a phase portrait is an important aid to understanding the system behaviour. The results can also be useful for comparing results obtained by the approximate DF method, to be introduced in the next chapters, when it is applied to second order systems.

## 2.5 References

- 2.1 Atherton, DP 1975, Nonlinear Control Engineering: Describing Function Analysis and Design, Chapter 2, Van Nostrand Reinhold, London.
- 2.2 Astrom, KJ & Murray, RM 2008, Feedback Systems: An Introduction for Scientists and Engineers, Chapters 2, 3 & 4 Princeton University Press.



## 2.6 Bibliography

Blaquiere, A 1966, Nonlinear Systems Analysis, Academic Press, New York.

Cosgriff, RL 1958, Nonlinear Control Systems, McGraw Hill, New York.

Cunningham, WJ 1958, Introduction to Nonlinear Analysis, McGraw Hill, New York.

Gelb, A & Vander Velde, WE 1968, Multiple-Input Describing Functions and Nonlinear System Design, McGraw Hill, New York.

Gibson, JE 1963, Nonlinear Automatic Control, McGraw Hill, New York.

Graham, D & McRuer, D 1961, Analysis of Nonlinear Control Systems, Wiley, New York.

Hayashi, C 1964, Nonlinear Oscillations in Physical Systems, McGraw Hill, New York.

Kalman, RE 1954, Phase plane analysis of automatic control systems with nonlinear gain elements. Trans. AIEE, Vol 73(II), pp 383-390.

Kalman, RE 1955, Analysis and design principles of second and higher order saturating servomechanisms. Trans. AIEE, Vol74(II), pp 294-308.

Minorsky, N 1962, Nonlinear Oscillations, Van Nostrand, New York.

Slotine, JJE & Li, W 1991, Applied Nonlinear Control, Prentice Hall, New Jersey.

Struble, RA 1962, Nonlinear Differential Equations, McGraw Hill, New York.

Thaler, GJ & Pastel, MP 1962, Analysis and Design of Nonlinear Feedback Control Systems, McGraw Hill, New York.

Van der Pol, B 1934, Nonlinear theory of electric oscillations. Proc. IRE, Vol 22, pp 1051-1086.

West, JC. 1960, Analytical Techniques for Nonlinear Control Systems, EUP, London.

West, JC, Douce, JL & 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, Vol 101, pp 156-165.

Please click the advert



### The financial industry needs a strong software platform

That's why we need you

SimCorp is a leading provider of software solutions for the financial industry. We work together to reach a common goal: to help our clients succeed by providing a strong, scalable IT platform that enables growth, while mitigating risk and reducing cost. At SimCorp, we value commitment and enable you to make the most of your ambitions and potential.

Are you among the best qualified in finance, economics, IT or mathematics?

Find your next challenge at  
[www.simcorp.com/careers](http://www.simcorp.com/careers)



[www.simcorp.com](http://www.simcorp.com)

MITIGATE RISK | REDUCE COST | ENABLE GROWTH



# 3 The Describing Function

## 3.1 Introduction

The describing function, which will be abbreviated DF, method was developed simultaneously in several countries during the 1940s. Engineers found that control systems which were being used in many applications, for example gun pointing and antenna control, could exhibit limit cycles under certain conditions rather than move to a static equilibrium. They realized this instability was due to nonlinearities, such as backlash in the gears of the control system, and they wished to obtain a design method which could ensure the resulting systems were free from limit cycle operation. They observed that when limit cycles occurred the waveforms at the system output were often approximately sinusoidal and this indicated to them a possible analytical approach, namely to assume that the signal at the input to the nonlinear element in the loop was a sinusoid. 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 they can be used to study. More will be said on these aspects later but we begin by considering the initial problem of investigating the possibility of a limit cycle in a feedback system using the DF or S (sinusoidal) DF as it is often named.

Consider the feedback system shown in Figure 3.1 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, that is with  $r(t) = 0$ , with the output  $c(t)$  approximately sinusoidal, then the input  $x(t)$  to the nonlinearity might also be expected to be near 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. This approach had previously been used by physicists to investigate such aspects as the generation of oscillations in electronic circuits; the Van der Pol oscillator mentioned in the previous chapter being an example. The DF of a nonlinearity is 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.

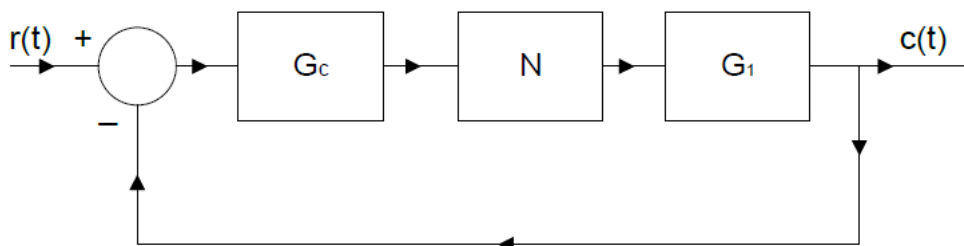


Figure 3.1 A simple nonlinear feedback system

## 3.2 The Sinusoidal Describing Function

Assume that in Figure 3.1,  $x(t)$  the input to the nonlinearity  $N$ , defined as  $n(x)$ , is  $x(t) = a \cos \theta$ , where  $\theta = \omega t$  and  $n(x)$  is a symmetrical odd nonlinearity, then the output  $y(t)$  will be given by the Fourier series.

$$y(\theta) = \sum_{n=0}^{\infty} a_n \cos n\theta + b_n \sin n\theta \quad (3.1)$$

where  $a_n = b_n = 0$  for  $n$  even, and in particular

$$a_1 = (1/\pi) \int_0^{2\pi} y(\theta) \cos \theta d\theta \quad (3.2)$$

and

$$b_1 = (1/\pi) \int_0^{2\pi} y(\theta) \sin \theta d\theta \quad (3.3)$$

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

$$N(a) = (a_1 - jb_1)/a \quad (3.4)$$

which may be written

$$N(a) = N_p(a) + jN_q(a) \quad (3.5)$$

where

$$N_p(a) = a_1/a \text{ and } N_q(a) = -b_1/a \quad (3.6)$$

Alternatively, in polar co-ordinates

$$N(a) = M(a)e^{j\psi(a)} \quad (3.7)$$

where

$$M(a) = (a_1^2 + b_1^2)^{1/2}/a \quad (3.8)$$

and

$$\psi(a) = -\tan^{-1}(b_1/a_1). \quad (3.9)$$

If  $n(x)$  is single valued it is easily shown that  $b_1 = 0$  and from symmetry considerations as  $n(x)$  is odd, that

$$a_1 = (4/\pi) \int_0^{\pi/2} y(\theta) \cos \theta d\theta \quad (3.10)$$

giving

$$N(a) = (a_1/a) = (4/a\pi) \int_0^{\pi/2} y(\theta) \cos \theta d\theta \quad (3.11)$$

Although equations (3.2) and (3.3) 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

$$a_1 = (2/a) \int_{-a}^a x n_p(x) p(x) dx \quad (3.12)$$

$$b_1 = (2/a\pi) \int_{-a}^a n_p(x) dx \quad (3.13)$$

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

$$p(x) = (1/\pi)(a^2 - x^2)^{-1/2} \quad (3.14)$$

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

$$n_p(x) = [n_1(x) + n_2(x)]/2 \quad (3.15)$$

and

$$n_q(x) = [n_2(x) - n_1(x)]/2 \quad (3.16)$$

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  $n_q(x) = 0$  and integrating equation (3.12) by parts gives

$$a_1 = (4/\pi)n(0^+) + (4/a\pi) \int_0^a n'(x)(a^2 - x^2)^{1/2} dx \quad (3.17)$$

where  $n'(x) = dn(x)/dx$  and  $n(0^+) = \lim_{\varepsilon \rightarrow 0} n(\varepsilon)$ ; a useful expression for obtaining DFs for linear segmented characteristics.

### 3.3 Some Properties of the DF

An advantage of the formulation of equations (3.12) and (3.13) 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:  

$$N_q(a) = - (1/a^2 \pi) (\text{area of nonlinearity loop})$$
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  

$$N_\alpha(a) < N_\beta(a) \text{ for input amplitudes } a \text{ less than } b.$$

3. For the sector bounded single-valued nonlinearity that is  $k_1x < n(x) < k_2x$  for all  $0 < x < b$  then  $k_1 < N(a) < k_2$  for 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.

It can also be shown that the gain  $K$  is the best transfer function model in the mean squared error sense. That is the optimum  $G(s)$  is  $K$ .

Please click the advert



What do you want to do?

No matter what you want out of your future career, an employer with a broad range of operations in a load of countries will always be the ticket. Working within the Volvo Group means more than 100,000 friends and colleagues in more than 185 countries all over the world. We offer graduates great career opportunities – check out the Career section at our web site [www.volvogroup.com](http://www.volvogroup.com). We look forward to getting to know you!

**VOLVO**  
AB Volvo (publ)  
[www.volvogroup.com](http://www.volvogroup.com)

VOLVO TRUCKS | RENAULT TRUCKS | MACK TRUCKS | VOLVO BUSES | VOLVO CONSTRUCTION EQUIPMENT | VOLVO PENTA | VOLVO AERO | VOLVO IT  
VOLVO FINANCIAL SERVICES | VOLVO 3P | VOLVO POWERTRAIN | VOLVO PARTS | VOLVO TECHNOLOGY | VOLVO LOGISTICS | BUSINESS AREA ASIA

Download free ebooks at [bookboon.com](http://bookboon.com)

### 3.4 The Evaluation of some DFs

To illustrate the evaluation of the DF for some specific nonlinearities a few simple examples are considered below.

#### 3.4.1 Cubic Nonlinearity

For this nonlinearity  $n(x) = x^3$  and using equation (3.10) one has,

$$\begin{aligned}
 a_1 &= (4/\pi) \int_0^{\pi/2} (a \cos \theta)^3 \cos \theta d\theta \\
 &= (4/\pi) a^3 \int_0^{\pi/2} \cos^4 \theta d\theta \\
 &= (4/\pi) a^3 \int_0^{\pi/2} \left( \frac{3}{8} + \frac{\cos 2\theta}{2} + \frac{\cos 4\theta}{8} \right) d\theta = 3a^3/4 \text{ giving} \\
 N(a) &= 3a^2/4
 \end{aligned} \tag{3.18}$$

Alternatively from equation (3.12) and using the fact that  $n(x)$  is odd, one has

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

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

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

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

#### 3.4.2 Saturation Nonlinearity

The DF can also be found by taking the nonlinearity input as  $a \sin \theta$ , in which case for the ideal saturation characteristic shown in Figure 3.2, with unit slope in the linear regime and saturation for an input above  $\delta > 0.5$ , the nonlinearity output waveform  $y(\theta)$  is as shown in the same figure.

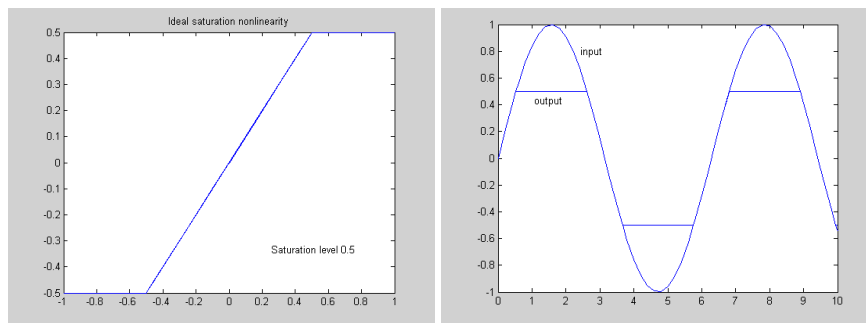


Figure 3.2 Saturation nonlinearity, input sinusoidal and output waveforms

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 for a linear regime slope of  $m$  and saturation at  $\delta$  gives

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

which for  $a > \delta$  gives

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

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

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

which on substituting for  $\delta$  gives the result

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

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

$$N_s(\delta/a) = \begin{cases} 1 & \text{for } a < \delta \\ (1/\pi)[2\alpha + \sin 2\alpha] & \text{for } a > \delta \end{cases} \quad (3.19)$$

Alternatively one can evaluate,  $N(a)$ , from equation (3.17), which yields, as the nonlinearity slope is zero for inputs greater than  $\delta$

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

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

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

### 3.4.3 Relay with Dead Zone and Hysteresis

The characteristic of a relay with dead zone and hysteresis, for  $h = \delta = 1$  and  $\Delta = 0.5$ , is shown in Figure 3.3 together with the corresponding input, assumed equal to  $a \cos \theta$ , and output waveforms.

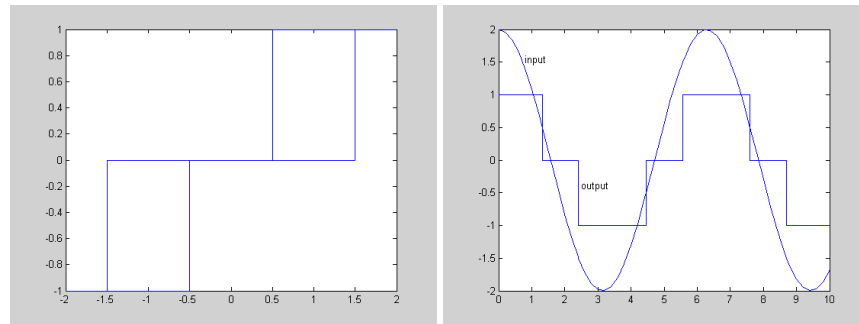


Figure 3.3 Relay with dead zone and hysteresis, and input and output waveforms

Using equations (3.2) and (3.3) over the interval  $-\pi/2$  to  $\pi/2$  and assuming that the input amplitude  $a$  is greater than  $\delta + \Delta$ , gives

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

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

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

$$N(a) = \frac{2h}{a^2 \pi} \{ [a^2 - (\delta + \Delta)^2]^{1/2} + [a^2 - (\delta - \Delta)^2]^{1/2} \} - \frac{j4h\Delta}{a^2 \pi} \quad (3.20)$$

Please click the advert

RedStarResume.

## Do you want your Dream Job?

More customers get their dream job by using RedStarResume than any other resume service.

RedStarResume can help you with your job application and CV.

Go to: [Redstarresume.com](http://Redstarresume.com)  
Use code "BOOKBOON" and save up to \$15

(enter the discount code in the "Discount Code Box")



For the alternative approach one must first obtain the in-phase and quadrature nonlinearities which are shown in Figure 3.4. Using equations (3.12) and (3.13) one obtains

$$a_1 = (4/a) \int_{\delta-\Delta}^{\delta+\Delta} x(h/2)p(x)dx + \int_{\delta+\Delta}^a xhp(x)dx$$

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

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

The DFs for other relay characteristics can easily be found from this result or worked out directly. The latter is very straightforward when the relay has no dead zone as its output is a square wave of amplitude  $\pm h$ , which has a fundamental component of amplitude  $4h/\pi$ , either in phase with the input for  $\Delta = 0$ , or lagging when  $\Delta$  is finite. Using equation 3.20 for no hysteresis,  $\Delta = 0$ ; for no dead zone,  $\delta = 0$ ; and for an ideal relay,  $\Delta = \delta = 0$ .

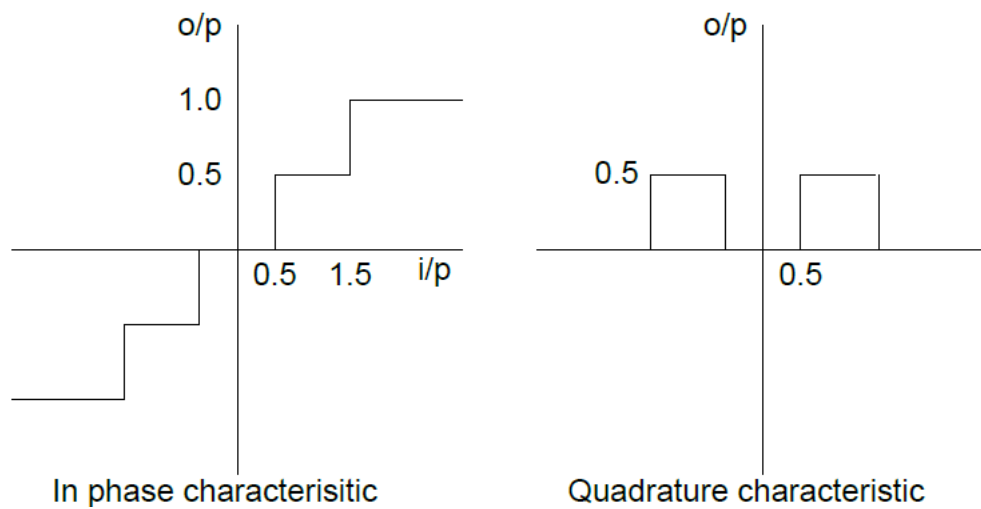


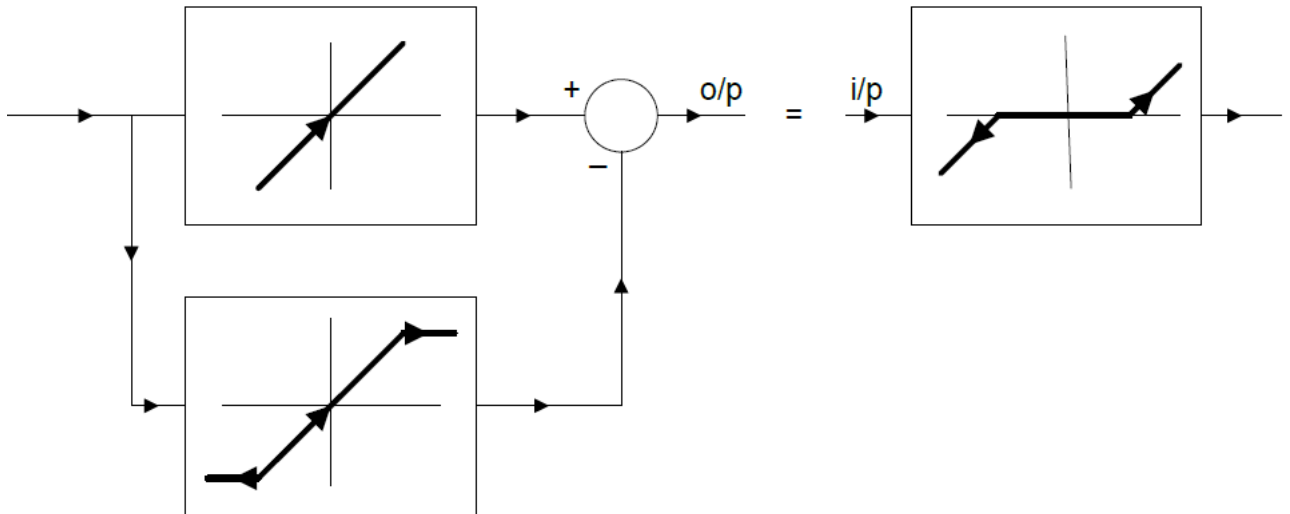
Figure 3.4 In phase and quadrature nonlinearities

It is possible to model a typical double valued nonlinearity like the above as a nonlinear function  $n(x, \dot{x})$  and the describing function theory based on this can be found in reference 3.1. Unfortunately this description is not correct for any input signal,  $x$ , and it has been used to obtain erroneous results for  $x$  a random signal. Two tables of DFs for a sinusoidal input are given in the Appendix 3.11, one for single valued nonlinearities and the other for double valued nonlinearities. Extensive use is made in the tables of the DF for saturation,  $N_s(\delta/a)$ , defined in equation (3.19)

### 3.5 Nonlinear Models and DFs

Nonlinear models, which consist of linear segments with multiple break points, can often be modeled from the nonlinearities in (b) and (c) above and linear gains, usually connected in parallel. For example, the dead zone characteristic which for a positive input has a zero output until a value of  $\delta$  and then a slope of unity is easily shown to be modeled by a unit gain in parallel with a saturation characteristic with slope  $m = -1$ , as illustrated in Figure 3.5. Also a quantized characteristic can be modeled from relay characteristics with dead zone and no hysteresis in parallel.





**Figure 3.5** Modelling of an ideal dead zone characteristic from a unit gain and ideal saturation

Since it is easily shown that the DF of two nonlinearities in parallel is equal to the sum of their individual DFs, the DFs of linear segmented characteristics with multiple break points can easily be written down from the DFs of simpler characteristics. 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 quantized characteristic, linear segmented characteristic or Fourier series.

### 3.6 Harmonic Outputs

It is sometimes of value to know the magnitudes of the other harmonic outputs from a nonlinearity with a sinusoidal input. These can be obtained from the usual Fourier series results for equation (3.1) of

$$a_n = (1/\pi) \int_0^{2\pi} n(x) \cos n\theta d\theta \quad (3.21)$$

and

$$b_n = (1/\pi) \int_0^{2\pi} n(x) \sin n\theta d\theta \quad (3.22)$$

In terms of the amplitude probability density function for a sinusoid it can be easily shown that these expressions become

$$a_n = 2 \int_{-a}^a n(x) T_n(x/a) p(x) dx \quad (3.23)$$

and

$$b_n = 2 \int_{-a}^a n(x) D_n(x/a) p(x) dx \quad (3.24)$$

where  $T_n(x/a) = \cos[n \cos^{-1}(x/a)]$  and  $D_n(x/a) = \sin[n \cos^{-1}(x/a)]$

Since for single valued nonlinearities  $b_n = 0$  the more important result is for  $a_n$ . Here  $T_n(x/a)$  is the Chebyshev polynomial of order  $n$  and its value for any  $n$  can easily be found from the recursion relationship

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \quad (3.25)$$

with  $T_0(x) = 1$  and  $T_1(x) = x$

### 3.7 Sine plus Bias DF and the IDF

So far only DFs for nonlinearities with odd symmetry have been considered. For many control engineering situations this may be satisfactory, as for example in the consideration of torque saturation in a motor. Many situations do occur, however, where nonlinearities do not have odd symmetry and also even when they do the operation may not be about the odd symmetric axis. Thus, to handle the general situation for a single valued nonlinearity with characteristic  $n(x)$  one needs to consider an input consisting of a sinusoid plus bias, that is  $\gamma + a \cos \theta$  and evaluate both the sinusoidal and bias output of the nonlinearity. The describing function for the nonlinearity, known as the SBDF, has two components  $N_\gamma(a, \gamma)$  and  $N_a(a, \gamma)$  the gains to the bias and sinusoid, respectively. It is easily shown following the procedures of section 3.2 that

$$N_\gamma(a, \gamma) = (a_0/\gamma) = (1/\gamma) \int_{-a}^a n(x + \gamma)p(x) dx \quad (3.26)$$

and

$$N_a(a, \gamma) = (a_1/a) = (2/a) \int_{-a}^a n(x + \gamma)(x/a)p(x) dx \quad (3.27)$$

Thus for the input  $\gamma + a \cos \theta$  the output is  $\gamma N_\gamma(\gamma, a) + N_a(\gamma, a)a \cos \theta$ . The important point about the SBDF model is the interaction between the signals caused by the nonlinearity, so that the gain to both signals is affected by a change in one of them. If  $a$  is constant and  $\gamma$  varied the bias output of equation (3.26) can be plotted against  $\gamma$  to give a new nonlinear characteristic. This is often called a modified nonlinearity and denoted by  $n(a, \gamma)$  rather than  $a_0$ . The slope of the modified nonlinearity at any input value  $\gamma$  gives the gain  $g(a, \gamma)$  to any small unrelated signal in the presence of  $a$  and  $\gamma$ . That is

$$g(a, \gamma) = \frac{d}{d\gamma} \int_{-a}^a n(x + \gamma)p(x) dx \quad (3.28)$$

In particular, the gain to a small unrelated signal in the presence of  $x$  alone is  $g(a, 0)$ , usually called the incremental describing function (IDF) and denoted by  $N_{i\gamma}(a)$ , and is given by

$$N_{i\gamma}(a) = g(a, 0) = \int_{-a}^a n'(x)p(x) dx \quad (3.29)$$

where  $n'(x) = dn(x)/dx$ . Equation (3.29) is the bias output from the nonlinearity  $n'(x)$  with input  $x$ . For any single valued nonlinearity with  $n(0^+) = 0$ , equation (3.17) gives

$$a^2 N(a) = (2/\pi) \int_{-a}^a n'(x)(a^2 - x^2)^{1/2} dx$$

which on differentiating both sides with respect to  $a$  gives

$$2aN(a) + a^2 dN(a)/da = 2 \int_{-a}^a an'(x)p(x)dx$$

The left hand side is  $2aN_{i\gamma}(a)$ , so that

$$N_{i\gamma}(a) = N(a) + (a/2)N'(a) \quad (3.30)$$

where  $N'(a) = dN(a)/da$ .

### 3.8 Conclusions

This chapter has covered the definition, some properties, and approaches to evaluating the DF of a nonlinearity for a sinusoidal input. The approach has then been extended to cover the situation of a sinusoidal plus bias input. The next chapter will show how these results can be used in the approximate analysis of a simple nonlinear feedback system.

### 3.9 References

3.1 Atherton, DP 1975, Nonlinear Control Engineering: Describing Function Analysis and Design, Van Nostrand Reinhold, London.

3.2 Gelb, A & Vander Velde, WE 1968, Multiple-Input Describing Functions and Nonlinear System Design, McGraw Hill, New York.

### 3.10 Bibliography

Atherton, DP 1975, Nonlinear Control Engineering: Describing Function Analysis and Design, Van Nostrand Reinhold, London.

Atherton, DP 1981, Stability of Nonlinear Systems, Research Studies Press, John Wiley, Chichester.

Cook, PA 1994, Nonlinear Dynamical Systems, 2<sup>nd</sup> ed. Prentice-Hall International, London.

Cosgriff, RL 1958, Nonlinear Control Systems, McGraw Hill, New York.

Gelb, A & Vander Velde, WE 1996, Multiple-Input Describing Functions and Nonlinear System Design. McGraw Hill, New York

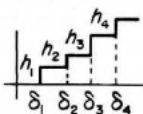
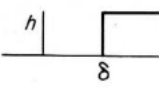
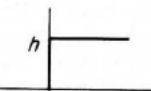
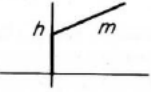
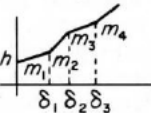
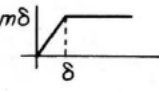
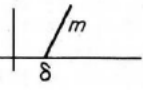
Gibson, JE 1963, Nonlinear Automatic Control, McGraw Hill, New York.

Graham, D & McRuer, D 1961, Analysis of Nonlinear Control Systems, Wiley, New York.

Slotine JJE & Li W 1991, Applied Nonlinear Control, Prentice-Hall International, New Jersey.

### 3.11 Appendix -Tables of Describing Functions

Two tables of DFs, or SDFs, are given below, one for single valued nonlinearities and the other for double valued nonlinearities. Extensive use is made in the tables of the describing function for saturation,  $N_s(\delta/a)$ , defined in eqn. (3.19). More comprehensive tables are given in other books [3.1, 3.2] for other describing functions, such as the SBDF and IDF of section 3.7, and for other inputs.

Nonlinearity	Comments	$N_p(a) = a_1/a$
General quantizer 	$a < \delta_1$ $\delta_{M+1} > a > \delta_M$	$N_p = 0$ $N_p = (4/a^2\pi) \sum_{m=1}^M h_m (a^2 - \delta_m^2)^{1/2}$
Uniform quantizer $h_1 = h_2 = \dots = h$ $\delta_m = (2m-1)\delta/2$	$a < \delta$ $(2M+1)\delta > a > (2M-1)\delta$ $n = (2m-1)/2$	$N_p = 0$ $N_p = (4h/a^2\pi) \sum_{m=1}^M (a^2 - n^2\delta^2)^{1/2}$
Relay with dead zone 	$a < \delta$ $a > \delta$	$N_p = 0$ $N_p = 4h(a^2 - \delta^2)^{1/2}/a^2\pi$
Ideal relay 		$N_p = 4h/a\pi$
Preload 		$N_p = (4h/a\pi) + m$
General piecewise linear 	$a < \delta_1$ $\delta_{M+1} > a > \delta_M$	$N_p = (4h/a\pi) + m_1$ $N_p = (4h/a\pi) + m_{M+1}$ $+ \sum_{j=1}^M (m_j - m_{j+1})N_s(\delta_j/a)$
Ideal saturation 		$N_p = mN_s(\delta/a)$
Dead zone 		$N_p = m[1 - N_s(\delta/a)]$

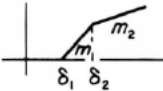
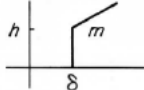
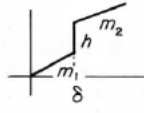
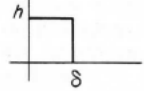
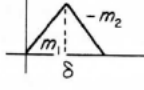
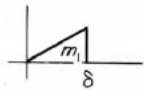
		$N_p = -m_1 N_s(\delta_1/a) + (m_1 - m_2) N_s(\delta_2/a) + m_2$
	$a < \delta$	$N_p = 0$
	$a > \delta$	$N_p = 4h(a^2 - \delta^2)^{1/2}/a^2\pi + m - mN_s(\delta/a)$
	$a < \delta$	$N_p = m_1$
	$a > \delta$	$N_p = (m_1 - m_2) N_s(\delta/a) + m_2 + 4h(a^2 - \delta^2)^{1/2}/a^2\pi$
	$a < \delta$	$N_p = 4h/a^2\pi$
	$a > \delta$	$N_p = 4h/[a - (a^2 - \delta^2)^{1/2}]/a^2\pi$
Limited field of view		
		$N_p = (m_1 + m_2) N_s(\delta/a) - m_2 N_s[(m_1 + m_2)\delta/m_2 a]$
	$a < \delta$	$N_p = m_1$
	$a > \delta$	$N_p = m_1 N_s(\delta/a) - 4m_1\delta(a^2 - \delta^2)^{1/2}/a^2\pi$
$y = x^m$	$m > -2$ $\Gamma$ is the gamma function	$N_p = \frac{\Gamma(m+1)a^{m-1}}{2^{m-1}\Gamma[(3+m)/2]\Gamma[(1+m)/2]} = \frac{2}{\sqrt{\pi}} \frac{\Gamma[(m+2)/2]a^{m-1}}{\Gamma[(m+3)/2]}$
$y = x^n$	$n$ odd integer. $\mu_n$ is the $n$ th moment of the amplitude distribution function of a sinusoid	$N_p = \frac{n(n-2)(n-4)\cdots 3}{(n+1)(n-1)\cdots 4} a^{n-1} = \frac{2}{a^2} \mu_{n+1}$
$y = x^n$	$n$ even integer	$N_p = \frac{4}{\pi} \frac{n(n-2)\cdots 2}{(n+1)(n-1)\cdots 3} a^{n-1}$
Harmonic nonlinearity $y = A_m \sin mx$	$J_1(ma)$ is the Bessel function of order 1	$N_p = 2A_m J_1(ma)/a$
$y = A_m \sinh mx$	$I_1(ma)$ is the modified Bessel function of order 1	$N_p = 2A_m I_1(ma)/a$
Exponential saturation $y = 1 - e^{-cx}$	$I_1(ca)$ is the modified Bessel function of order 1. $S_1(ca)$ is the modified Struve function of order 1	$N_p = (2/a)[I_1(ca) - S_1(ca)]$
$y = \frac{cx}{(1 + c^2 x^2)}$	$K(k)$ and $E(k)$ are complete elliptic integrals $u = ca/(1 + c^2 a^2)^{1/2}$	$N_p = (4/\pi)[(-u/c^2 a^3)K(u) + (1/au)E(u)]$

Table 3.1 Describing Functions of Single Valued Nonlinearities

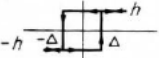
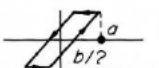
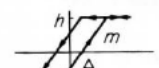
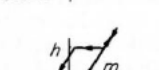
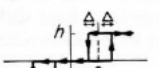
Nonlinearity	Comments	$N_p(a) = a_1/a$	$N_q(a) = -b_1/a$
On-off relay with hysteresis 	$a < \Delta$	$N_p = 0$	$N_q = 0$
	$a > \Delta$	$N_p = 4h(a^2 - \Delta^2)^{1/2}/a^2\pi$	$N_q = -4h\Delta/a^2\pi$
Friction-controlled backlash   	$a > b/2$	$N_p = \{1 + N_s[(a - b)/a]\}/2$	$N_q = b(b - 2a)a^2\pi$
	$a > \Delta + (h/m)$	$N_p = \frac{m}{2} \left[ N_s \left( \frac{h - m\Delta}{ma} \right) + N_s \left( \frac{h + m\Delta}{ma} \right) \right]$	$N_q = -4h\Delta/a^2\pi$
	$a > \Delta + (h/m)$	$N_p = \frac{m}{2} \left[ 2 - N_s \left( \frac{h + m\Delta}{ma} \right) + N_s \left( \frac{h - m\Delta}{ma} \right) \right]$	$N_q = -4h\Delta/a^2\pi$
	$m_1 > m_2$ $a > \frac{h}{m_1} + \frac{\Delta m_1}{(m_1 - m_2)}$ $a = \frac{\Delta m_1}{(m_1 - m_2)}$	$N_p = \left( \frac{m_1 - m_2}{2} \right) \left[ N_s \left( \frac{h + m_1 a}{m_1 a} \right) + N_s \left( \frac{h - m_1 a}{m_1 a} \right) \right] + m_2$	$N_q = -4h\Delta/a^2\pi$
On-off relay with dead zone and hysteresis 	$a < \delta + \Delta$	$N_p = 0$	$N_q = 0$
	$a > \delta + \Delta$	$N_p = \frac{2h}{a^2\pi} \{ [a^2 - (\delta + \Delta)^2]^{1/2} + [a^2 - (\delta - \Delta)^2]^{1/2} \}$	$N_q = -4h\Delta/a^2\pi$

Table 3.2 Describing Functions of Double Valued Nonlinearities

# 4 Stability and Limit Cycles using the DF

## 4.1 Introduction

The problem of determining whether a linear feedback system is stable is simply one of determination of the roots of its characteristic equation. This can be done in several ways with a particularly elegant and convenient engineering one, since it can be done from knowledge of the open loop frequency response locus, being due to Nyquist. On the other hand the problem of determining precisely the stability of an autonomous nonlinear feedback system, even one simply containing one linear dynamic element and one nonlinearity, has not been solved, even though it has exercised the minds of mathematicians and engineers for many years. The DF provides an approximate method for answering the problem and will be addressed in this chapter, whilst some exact approaches, known as absolute stability methods, will be considered in chapter 8. The discussions will be restricted to the feedback loop structure of Figure 3.1. Since our initial investigations are concerned with the stability of the autonomous system, that is  $r(t) = 0$ , the two linear dynamic blocks are in series and can be represented by the single transfer function  $G(s)$ . Further the position of the single static nonlinearity although assumed in the forward path could equally well be in the feedback path.

**Try this...**



The sequence 2, 4, 6, 8, 10, 12, 14, 16, ... is the sequence of even whole numbers. The 100th place in this sequence is the number...?

Challenging? Not challenging? Try more >>

[www.alloptions.nl/life](http://www.alloptions.nl/life)

Please click the advert

An early contribution to the problem was a conjecture by Aizermann. He conjectured that if a symmetrical odd nonlinearity was confined within a sector defined by straight lines of slope  $k_1$  and  $k_2$  then any nonlinear system with a nonlinearity lying entirely within the sector would be stable provided the linear system was stable for gains between  $k_1$  and  $k_2$ . Unfortunately the conjecture is not true, but it is if it is modified so that the last phrase reads ‘would be stable or possess a limit cycle provided the linear system was stable for gains between  $k_1$  and  $k_2$ ’. Thus, since use of the DF provides an approximate method for the determination of limit cycles, it also provides an approximate stability test for our simple feedback loop.

## 4.2 Limit Cycle Evaluation

To study the possibility of limit cycles in the autonomous closed loop system of Figure 3.1, the nonlinearity  $n(x)$  is replaced by its DF,  $N(a)$ . Thus, the open loop gain to a sinusoid is  $N(a)G(j\omega)$  and a limit cycle will exist if

$$N(a)G(j\omega) = -1 \quad (4.1)$$

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 assuming its passage through the nonlinearity is accurately described by  $N(a)$ . Since  $G(j\omega)$  is a complex function of  $\omega$  and  $N(a)$  may be a complex function of  $a$ , a solution to equation (4.1) will yield both the frequency  $\omega$  and amplitude  $a$  of an assumed sinusoidal limit cycle.

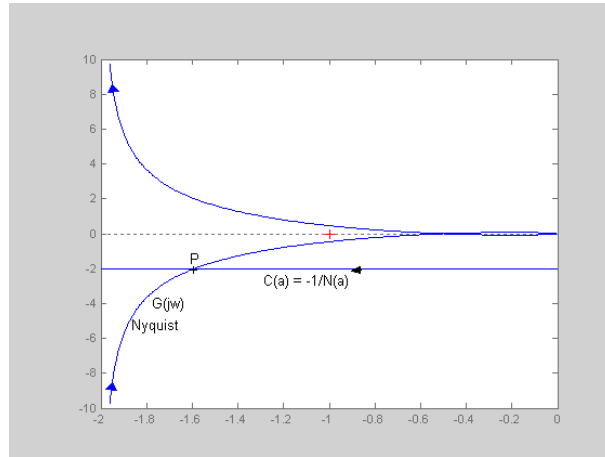
Various approaches can be used to examine equation (4.1) with the choice affected to some extent by the problem. For example, relevant factors might be whether the nonlinearity is single or double valued or whether  $G(j\omega)$  is available from a transfer function  $G(s)$  or as measured frequency response data. Typically the functions  $G(j\omega)$  and  $N(a)$  are plotted separately on Bode, Nyquist, or Nichols diagrams. Alternatively, stability criteria such as the Hurwitz-Routh or root locus plots may be used for the characteristic equation

$$1 + N(a)G(s) = 0 \quad (4.2)$$

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

Figure 4.1 illustrates the procedure on a Nyquist diagram, where the  $G(j\omega)$  and  $C(a) = -1/N(a)$  loci are plotted and shown intersecting at P for  $a = a_0$  and  $\omega = \omega_0$ . The DF method therefore indicates that the system has a limit cycle with the input sinusoid to the nonlinearity,  $x$ , equal to  $a_0 \sin(\omega_0 t + \phi)$ , where  $\phi$  depends on the initial conditions and time origin. In practice the limit cycle will not be sinusoidal and  $a_0$  is an approximation for the amplitude of the fundamental component of the limit cycle. Thus if one wishes to estimate the accuracy of the DF prediction for a limit cycle one should measure the amplitude of the fundamental not the peak amplitude of the waveform as is often done for convenience. 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.





**Figure 4.1** Limit cycle found from intersection of  $G(j\omega)$  and  $C(a)$  on a Nyquist plot

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

### 4.3 Stability of a Predicted Limit Cycle

A further point of interest when a solution to equation (4.1) exists is whether the predicted limit cycle is stable. This is obviously important if the control system is designed to have a limit cycle operation, as in the case of an on-off temperature control system, but it may also be important in other systems. If, for example, an unstable limit cycle condition is reached the signal amplitudes will not become bounded but continue to grow. The stability of a limit cycle, provided only one solution is predicted, can in most cases 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). This result is known as the Loeb intersection criterion and in terms of the approximate DF analysis, is a necessary, but not a sufficient criterion for a stable limit cycle.

To investigate the stability of a limit cycle one has to examine the perturbation equation. Consider the transfer function to be the ratio of two polynomials so that it can be written as

$$G(s) = p(s)/q(s) \quad (4.3)$$

Then if  $D$  denotes the differential operator the differential equation for the autonomous system can be written

$$q(D)x(t) + p(D)n(x(t)) = 0 \quad (4.4)$$

If the system has a limit cycle solution  $x^*(t)$  and this is perturbed by a small amount  $\delta x(t)$  then

$$q(D)[x^*(t) + \delta x(t)] + p(D)n(x^*(t) + \delta x(t)) = 0$$

This can be written as

$$q(D)[x^*(t) + \delta x(t)] + p(D)n(x^*(t)) + p(D)n'(x^*(t))\delta x(t) = 0$$

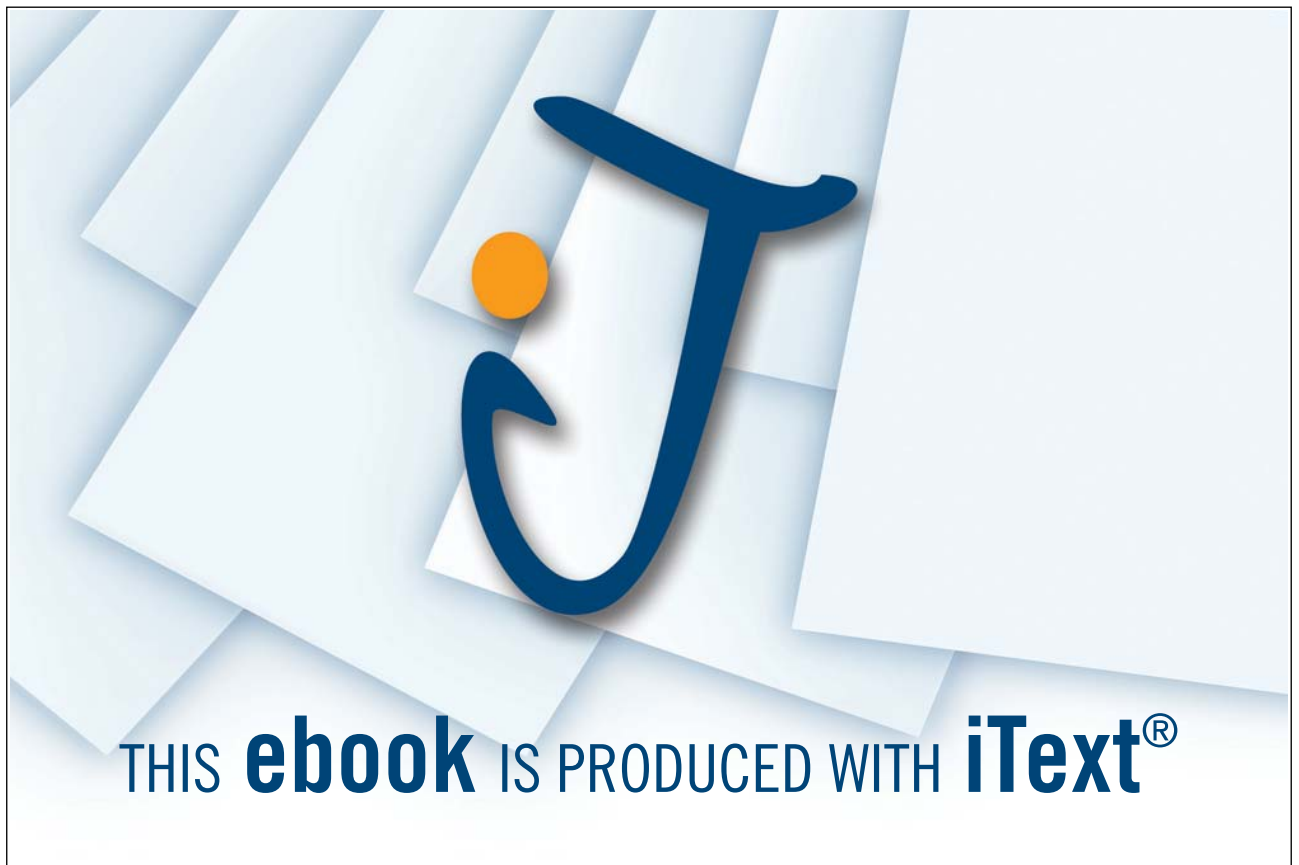
Thus on subtracting equation (4.4) one has the perturbation equation

$$q(D)\delta x(t) + p(D)n'(x^*(t))\delta x(t) = 0 \quad (4.5)$$

This is a time varying equation for the perturbation,  $\delta x(t)$ , which must be stable for the limit cycle to be stable. It can be shown that a necessary condition for this equation to be stable is that the time averaged equation must be stable. Then if one uses the approximate describing function solution for  $x^*(t)$ , the time averaged value of the nonlinear term is the IDF of equation (3.29), so that a necessary condition for stability of the limit cycle is that the equation  $q(D) + p(D)N_{ir}(a) = 0$  is stable. In terms of the system transfer function this is the characteristic equation

$$1 + N_{ir}(a)G(s) = 0 \quad (4.6)$$

Please click the advert



## 4.4 DF Accuracy

A specification for many control systems is that when perturbed by either a step input or disturbance they should return to 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 using the sector bounds of the nonlinearity provided any input does not cause a bias at the nonlinearity input 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 signal fed back to the nonlinearity, assuming the loop open at the nonlinearity input. If the percentage distortion in this signal is less than 5% then the DF prediction should be valid. It has also been shown theoretically that the estimate for the limit cycle frequency,  $\omega$ , is more accurate than for the amplitude,  $a$ . Further, as is often not clearly pointed out,  $a$  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  $\omega$  corresponding to the near intersection values of  $C(a) = G(j\omega)$ . Using the distortion criteria given above a very low value must exist for the prediction to be valid when the loci are very close. More details on these concepts are presented in section 5.5.

## 4.5 Some Examples of Limit Cycle Evaluation

Some 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. The first example taken is the Van der Pol equation discussed in section 2.2, although a first harmonic balance approach rather than the DF is used as the nonlinear term involves two ‘inputs’,  $x$  and  $\dot{x}$ . It is possible, however, to use DF theory for a nonlinear element defined by  $n(x, \dot{x})$  and the interested reader is referred to the first source in the bibliography. The second and third examples consider limit cycles in relay systems. It is shown in chapter 6 how limit cycles in relay systems can be calculated accurately so that comparisons of the results allows some comments to be made on DF accuracy. The final example is introduced to show the value of the IDF in determining limit cycle stability.

### 4.5.1 The Van der Pol Equation

If a harmonic balance approach is used for equation (2.4) one may assume  $x_1 = a \sin \omega t$  and  $x_2 = \dot{x}_1 = a\omega \cos \omega t$  and the equation can be written

$$\dot{x}_2 + x_1 = \mu(1 - x_1^2)x_2$$

Substituting for  $x_1$  and  $x_2$  and neglecting frequencies higher than  $\omega$  in the right hand side gives

$$-a\omega^2 \sin \omega t + a \sin \omega t = \mu(a\omega \cos \omega t - \frac{a^3 \omega}{4} \cos \omega t).$$

For this to be true requires  $\omega = 1$  for the left hand side and  $a\omega = a^3\omega/a$ , giving  $a = 2$ , for the right hand side. Thus, a limit cycle solution, which is independent of  $\mu$ , with frequency 1 rad/s and amplitudes of 2 for  $x_1$  and  $x_2$ , is predicted. The result is therefore only reasonably accurate for small values of  $\mu$ , where the phase plane plot is almost elliptical.

#### 4.5.2 Feedback Loop Containing a Relay with Dead Zone

For this example the feedback loop of Figure 4.1 is considered with  $n(x)$  a relay with dead zone and  $G(s) = 2/s(s+1)^2$ . The DF for this relay is obtained from equation (3.20) with  $\Delta = 0$  and is given by

$$N(a) = 4h(a^2 - \delta^2)^{1/2}/a^2\pi \text{ for } a > \delta,$$

which is real because the nonlinearity is single valued. A graph of  $N(a)$  against  $a$  is given in Figure 4.2.

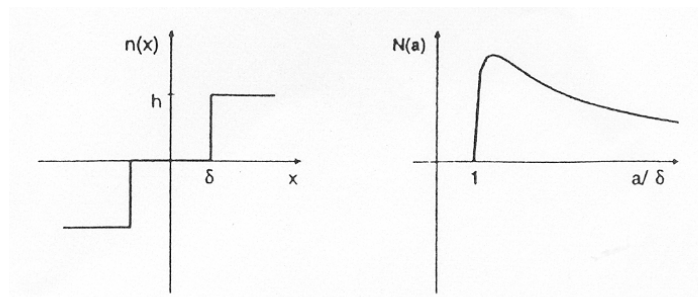
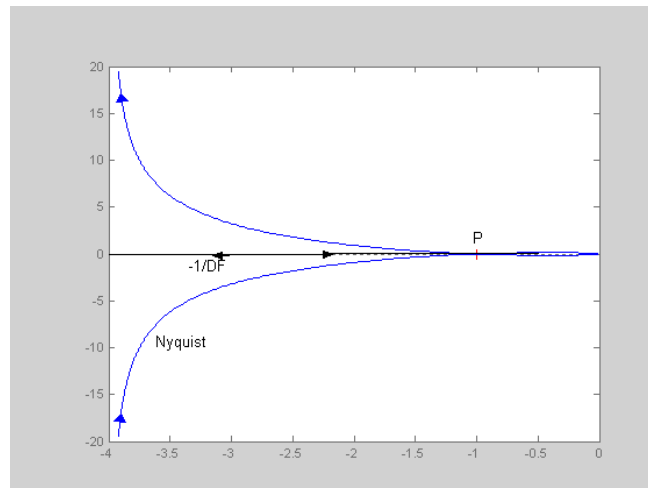


Figure 4.2 Relay with dead zone and its DF

It 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 Figure 4.3, lies on the negative real axis starting at  $-\infty$  and returning there after reaching a maximum value of  $-\pi\delta/2h$ . The given transfer function  $G(j\omega)$  crosses the negative real axis, as shown in Figure 4.3, at the point  $(-1, 0)$  at a frequency of  $\tan^{-1}\omega = 45^\circ$  that is  $\omega = 1$  rad/sec, and therefore cuts the  $C(a)$  locus twice at the point P. The two possible limit cycle amplitudes at this frequency can be found by solving

$$\frac{a^2\pi}{4h(a^2 - \delta^2)^{1/2}} = G(j\omega)|_{\omega=1} = 1$$

which assuming  $\delta = 1$  and  $h = \pi$  gives  $a = 1.04$  and  $3.86$ . Using either the Loeb or the IDF criterion shows the smaller amplitude limit cycle to be unstable and the larger one to be 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.

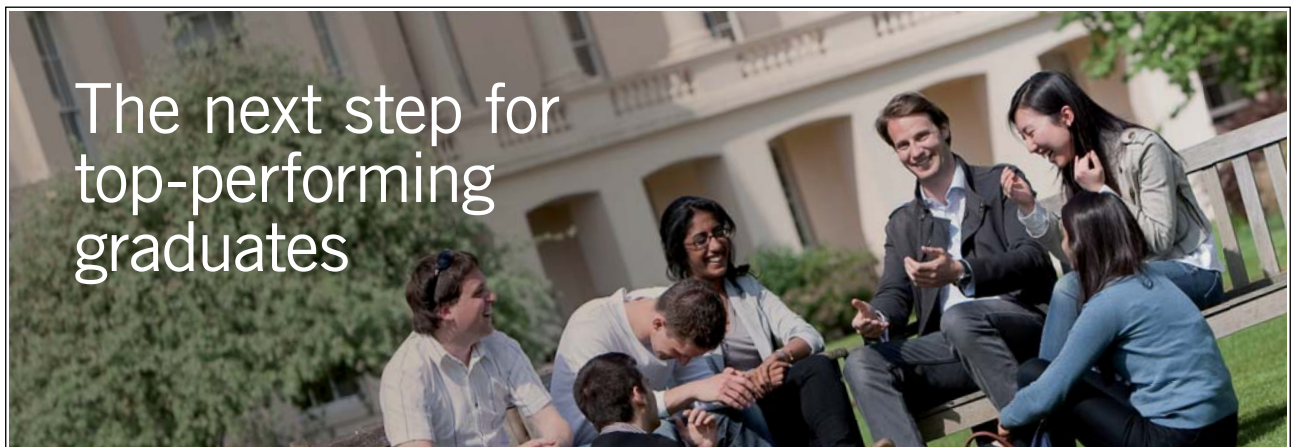


**Figure 4.3** Solutions for the two limit cycles at P one unstable the other stable

The exact frequencies of the limit cycles, obtained using the approach given in Chapter 6, 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 smaller amplitude (unstable) limit cycle is not predicted accurately because the output from the relay is a waveform with narrow pulses, which is 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 according to the DF method if

$$\left. \frac{K}{\omega(1+\omega^2)} \right|_{\omega=1} < \frac{\pi d}{2h} \quad (4.7)$$

Please click the advert



### Masters in Management



Designed for high-achieving graduates across all disciplines, London Business School's Masters in Management provides specific and tangible foundations for a successful career in business.

This 12-month, full-time programme is a business qualification with impact. In 2010, our MiM employment rate was 95% within 3 months of graduation\*; the majority of graduates choosing to work in consulting or financial services.

As well as a renowned qualification from a world-class business school, you also gain access to the School's network of more than 34,000 global alumni – a community that offers support and opportunities throughout your career.

For more information visit [www.london.edu/mm](http://www.london.edu/mm), email [mim@london.edu](mailto:mim@london.edu) or give us a call on **+44 (0)20 7000 7573**.

\* Figures taken from London Business School's Masters in Management 2010 employment report

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

### 4.5.3 Feedback Loop with On Off Relay

Consider the feedback loop of Figure 3.1 with  $n(x)$  an ideal relay,  $G_c(s) = 1$ , and the plant with a transfer function  $G_1(s) = 10/(s+1)^3$ . Equation (3.20) with  $\delta = \Delta = 0$  gives  $N(a) = 4h/a\pi$ . Thus the  $C(a)$  locus,  $-1/N(a) = -a\pi/4h$  and lies on the negative real axis starting from the origin with  $a = 0$ . The Nyquist locus  $G(j\omega)$  intersects it where it has a phase of  $-180^\circ$ . The values of  $a$  and  $\omega$  at the intersection can be calculated from

$$-a\pi/4h = \frac{10}{(1+j\omega)^3} \text{ which can be written}$$

$$\arg \frac{10}{(1+j\omega)^3} = 180^\circ \quad (4.8)$$

and

$$\frac{a\pi}{4h} = \frac{10}{(1+\omega^2)^{3/2}}. \quad (4.9)$$

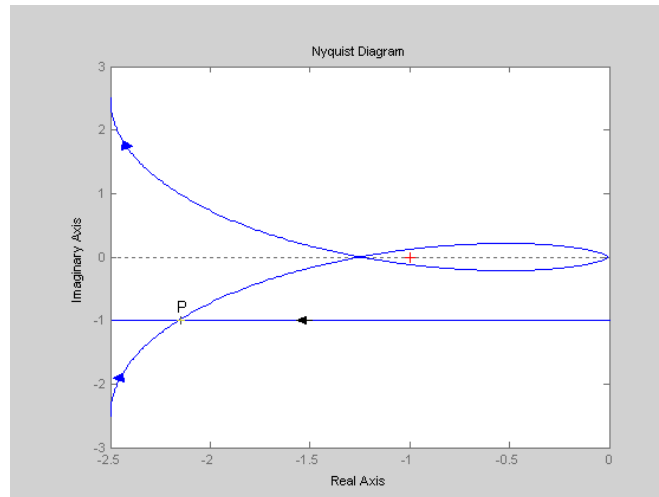
The solution for  $\omega_0$  from equation (4.8) is  $\tan^{-1}\omega_0 = 60^\circ$ , giving  $\omega_0 = \sqrt{3} = 1.732$ . Because the DF solution is approximate, the actual measured frequency of oscillation will differ from this value by an amount which will be smaller the closer the oscillation is to a sinusoid. The exact frequency of oscillation is 1.708 rads/s so that  $\omega_0$  is in error by 1.41%. A symmetrical square wave has a harmonic content which is  $1/n^{\text{th}}$  that of the fundamental for  $n$  odd. It is thus easy to calculate the harmonic components at the output of  $G(s)$  and the third and fifth are 5.40% and 1.21% of the fundamental. Solving equation (4.9) gives the amplitude of the assumed sinusoidal limit cycle  $a$  as  $5h/\pi$ .

If the relay has an hysteresis of  $\Delta$ , then putting  $\delta = 0$  in equation (3.20) gives

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

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

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 Figure 4.4 for  $\Delta = 1$  and  $h = \pi/4$  giving  $C(a) = -(a^2 - 1)^{1/2} - j$ . If the same transfer function as above 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  $a = 1.91$ .



**Figure 4.4** Nyquist plot for  $10/(1+s)^3$  and DF for relay with hysteresis

For the ideal relay, equation (3.30) gives  $N_{ir}(a) = 2h/a\pi$  for the IDF so that the roots of the characteristic equation  $1 + N_{ir}(a)G(s) = 0$  show the limit cycle is stable. This agrees with the perturbation approach which also shows that the limit cycle is stable when the relay has hysteresis.

#### 4.5.4 Use of the IDF

Consider an unstable plant with the transfer function

$$G(s) = K/(s - b)(s^2 + 2s + 10) \quad (4.10)$$

in a feedback loop with an ideal relay with output  $\pm 1$ . The characteristic equation for the loop with the relay gain taken as its DF gives

$$s^3 + s^2(2 - b) + s(10 - 2b) + k - 10b = 0 \quad (4.11)$$

where  $KN(a)$  is replaced by  $k$ . If a third order characteristic equation has a limit cycle at a frequency  $\omega$ , then its form will be

$$(s^2 + \omega^2)(s + c) = 0 \quad (4.12)$$

which gives

$(2 - b)(10 - 2b) = k - 10b$ , so that  $k = 20 - 4b + 2b^2$ . The frequency of the limit cycle is  $(10 - 2b)^{1/2}$  and the real root is at  $-c$  where  $-c = b - 2$ , which will be negative for  $b < 2$ . The situation is illustrated in Figure 4.5, which is a root locus plot for the transfer function of equation (4.10), with  $K = b = 1$ . However, when the limit cycle is established the gain through the relay to any unrelated signal is the IDF, which for the ideal relay is easily shown from equation (3.30) to be  $N(a)/2$ , that is it is halved and the new characteristic equation is

$$s^3 + s^2(2 - b) + s(10 - 2b) + 10 - 2b + b^2 - 10b = 0 \quad (4.13)$$

This will have a complex pair of roots and a real root and for stability of the limit cycle this latter root must be negative, which will require the constant in equation (4.13) to be negative. This will be the case for  $b < 0.9$ . Thus, the approximate DF theory indicates that the odd symmetrical limit cycle cannot exist for  $b > 0.9$ .

This was confirmed by simulation results which gave an odd symmetrical limit cycle to around  $b = 0.9$  and an asymmetrical limit cycle for values of  $b$  from around 0.9 to 1.3 at which point instability occurred, with the output from the relay locking at one of its limit levels of  $\pm 1$ .

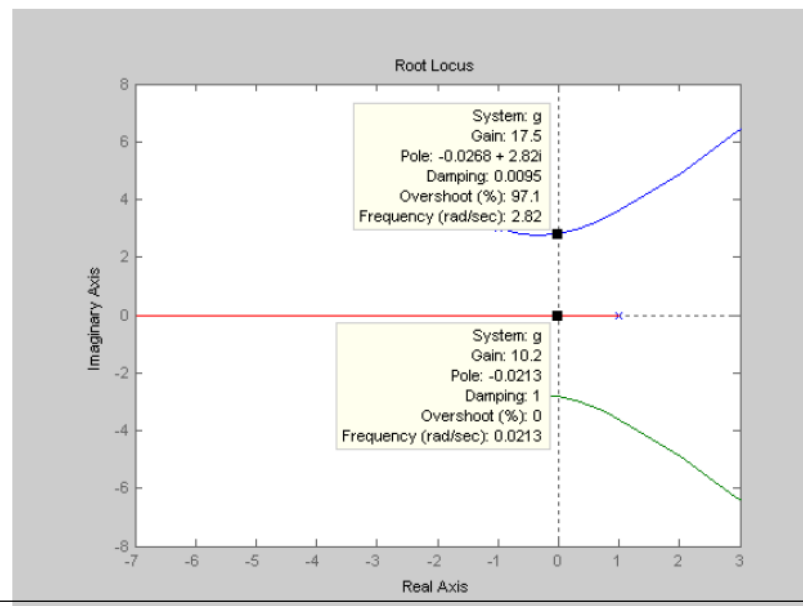


Figure 4.5 Root locus plot for transfer function of equation (4.10) for  $K = b = 1$ .



The m  
of the  
value,

ximate gains on the root locus for the parameters  $K = b = 1$ . The exact value  
he imaginary axis at  $\pm 2.82$  is 18 and when the gain is 9, half the previous  
ne at 0.123, indicating that an odd symmetrical limit cycle is unstable.

*You're full of energy  
and ideas. And that's  
just what we are looking for.*

Looking for a career where your ideas could really make a difference? UBS's Graduate Programme and internships are a chance for you to experience for yourself what it's like to be part of a global team that rewards your input and believes in succeeding together.

Wherever you are in your academic career, make your future a part of ours by visiting [www.ubs.com/graduates](http://www.ubs.com/graduates).

[www.ubs.com/graduates](http://www.ubs.com/graduates)



Please click the advert

© UBS 2010. All rights reserved.



It can be seen from the gains marked that the complex poles cross the imaginary axis for a gain of 18 and for that value of gain the real pole is at -1. However for a gain of 9 (half of 18) the real pole will be in the rhs of the  $s$  plane so that the system is unstable. If the plot is done for  $b = 0.9$  then the gain when the complex roots cross the axis will be twice that for the real root crossing the axis. The difficulty of achieving an odd symmetrical limit cycle for  $b$  slightly less than 0.9 is presumably due to errors in the DF and the small range of attraction of the limit cycle, which is a closed curve in three space.

## 4.6 More than one Nonlinear Element

When the closed loop system has more than one nonlinear element it is possible to write down nonlinear algebraic equations based on the DF method to give any possible limit cycle solutions. The difficulty with this approach, however, is that satisfactory accuracy can only be obtained if, when a limit cycle occurs, it is reasonable to assume the inputs at all the nonlinear elements are nearly sinusoidal. In the classical multivariable version of Figure 3.1 where  $G(s)$  is a transfer matrix and  $n(x)$  a matrix with nonlinear elements on the diagonal this may be reasonable since typical transfer function elements of  $G(s)$  will be low pass, which filter the higher harmonic content. However, one has to assume that the limit cycle is at a single frequency, which is normally the case in a multivariable control system feedback loop. A simple counterexample is easily obtained by lightly coupling two single loops which have limit cycles at two different frequencies. Dependent on various factors the loops may continue to maintain both frequencies or lock to a limit cycle with two related, say 1 to 3, dominant frequencies.

In a single loop configuration the assumption of sinusoidal signals into all the nonlinear elements is usually not justifiable. If, for example, another nonlinear element is included in Figure 3.1 prior to  $G_c(s)$  then it may be reasonable to assume this has a sinusoidal input but since most controller transfer functions  $G_c(s)$  are not low pass filters the signal into the nonlinearity before the plant may be quite highly distorted. Thus, if a DF analysis is done the results may be incorrect. A well known example of this is shown in the Simulink diagram of Figure 4.6 where both nonlinear elements are ideal saturations with limit levels  $\pm 1$  and the linear elements are a phase lead with a transfer function  $(1 + s)/(1 + 0.1s)$  and a plant with a transfer function of  $K/s(1 + s)^2$  with  $K = 9$ . The linear system and the nonlinear system without the saturation characteristic before the phase lead are both stable for  $K < 11$ . However the system of Figure 4.6 possesses a limit cycle for large initial conditions on the integrator.

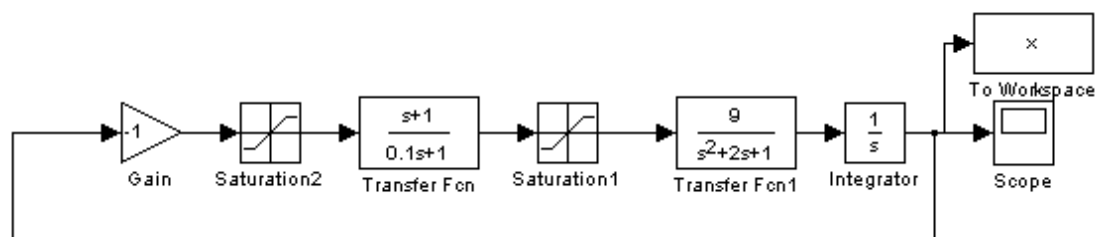
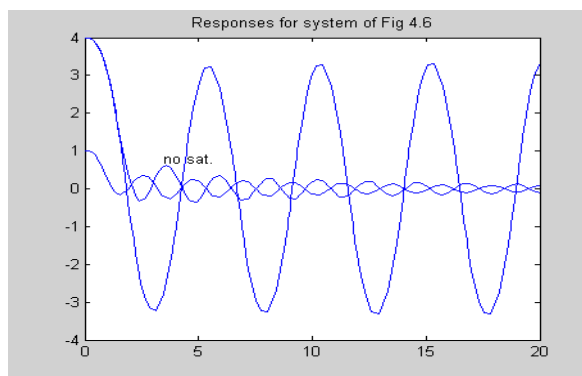
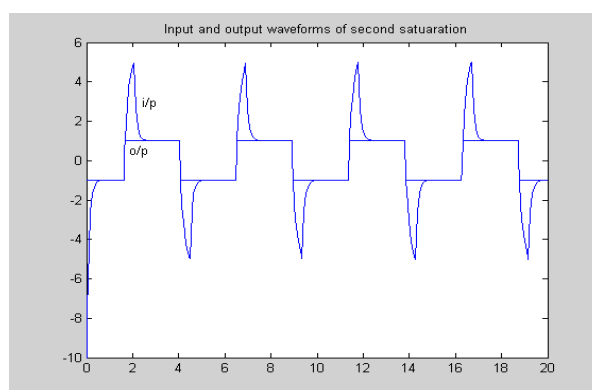


Figure 4.6 Simulink diagram of system with two ideal saturations



**Figure 4.7** Initial condition responses of system of Figure 4.6

Figure 4.7 shows the responses of the system of Figure 4.6 from initial conditions on the output integrator of 4 and 1 respectively. A limit cycle results from the former whereas for the second case the response is stable although very oscillatory. The third response shown on the figure labeled 'no sat.' is again for an initial condition of 4 but without the saturation before the lead network and is again very oscillatory. The limit cycle results because although it is almost sinusoidal at the integrator output it is highly distorted at the input to the second saturation, with the result that the fundamental component undergoes a phase shift when passing through the second saturation. Figure 4.8 shows the input and output waveforms for the limit cycle at the second saturation.



**Figure 4.8** Waveforms of limit cycle at the input and output of the second saturation

If with more than one nonlinearity in the loop it is reasonable to assume any limit cycle in the system would be approximately sinusoidal at the plant input then, as mentioned in section 1.3, if the plant consists of linear and nonlinear elements a frequency response model  $N(a, \omega)$  could be used to model it. This could be done either from mathematical modeling or experimental data, where if the distortion is small the output is taken as the signal at the same frequency as the sinusoidal input. This frequency dependent DF representation may be used in analysis and design, a topic discussed in section 9.5.

## 4.7 Applications of SBDF to find Limit Cycles

In this section some simple examples are considered to show the use of the SBDF in evaluating limit cycles. The first example considers a system with an asymmetrical nonlinearity in an autonomous feedback loop and the second an odd symmetrical nonlinearity in a loop with an input which causes a biased operation on the nonlinearity.

The procedure is to set up two equations, one for the first harmonic and the other for the bias, or dc, signal. The former is always the same, namely

$$N_a(a, \gamma)G(j\omega) = -1 \quad (4.14)$$

but the latter depends upon where any bias signal enters the loop and the value of  $G(0)$ . Since for a single valued nonlinearity  $N_a(a, \gamma)$  is real, the limit cycle frequency is given by

$$\arg G(j\omega) = -180^\circ \quad (4.15)$$

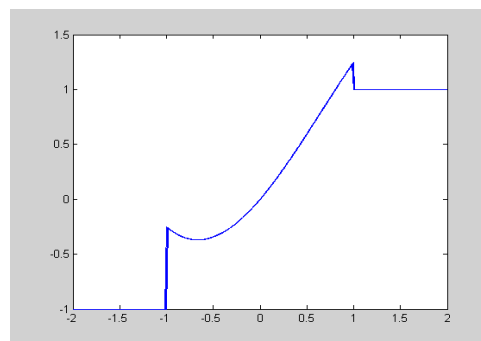
which if this frequency is  $\omega_c$ , gives

$$N_a(a, \gamma) = 1/|G(j\omega_c)| \quad (4.16)$$

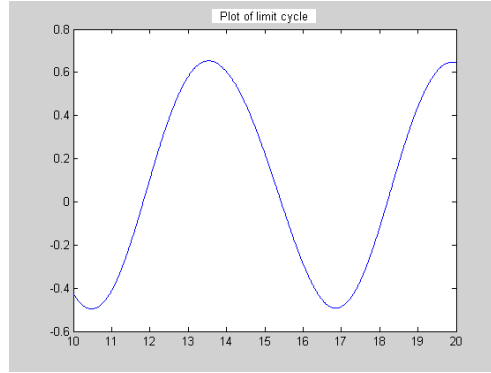
The frequency of the limit cycle given by the SBDF solution is thus, unlike the amplitude, unaffected by the bias signal, which will not normally be true in practice as the harmonic distortion will be affected by  $\gamma$

#### 4.7.1 Asymmetrical nonlinearity

The feedback system of Figure 3.1 is considered with no input,  $G(s) = 2.3/s(s+1)^2$  and the nonlinearity shown in Figure 4.9. It is defined by  $x + 0.5x^2 + 0.25x^3$  for  $|x| < 1$ , 1 for  $x \geq 1$  and -1 for  $x \leq -1$ . Figure 4.10 shows the resulting limit cycle at the output of  $G(s)$  from an output initial condition of 0.8 which since it lies in the range  $|x| < 1$ , it may be evaluated using its SBDF for this range.



**Figure 4.9** Nonlinear characteristic for Example of section 4.7.2



**Figure 4.10** Limit cycle in Example of section 4.7.1

For the polynomial nonlinearity  $n(x) = x^n$  the components of the SBDF from equations (3.26) and (3.27) are

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

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

respectively. After expanding  $(x + \gamma)^n$  it can be shown that

$$N_\gamma(a, \gamma) = \sum_{k=0}^n {}_nC_k \mu_k \gamma^{n-k-1} \quad (4.16)$$

$$N_a(a, \gamma) = (2/a^2) \sum_{k=0}^n {}_nC_k \mu_{k+1} \gamma^{n-k} \quad (4.17)$$

where the binomial coefficient

$${}_nC_k = \frac{n!}{(n-k)!k!} \quad (4.18)$$

Thus for the nonlinearity  $x + 0.5x^2 - 0.25x^3$  the values are

$$N_\gamma(a, \gamma) = 1 + 0.5\gamma + (a^2/4\gamma) - (3a^2/8) - (\gamma^2/4) \quad (4.19)$$

$$N_a(a, \gamma) = 1 + \gamma - (0.75a^2/4) - 0.75\gamma^2 \quad (4.20)$$

The linear transfer function has a phase shift of  $-180^\circ$  when  $\omega = 1$  with a corresponding gain of 1.15, so that using equation (4.16) gives

$$N_a(a, \gamma) = 1/1.15 = 0.87 \quad (4.21)$$

and since in the steady state the average value of the input to the integrator must be zero the bias equation is

$$\gamma N_\gamma(a, \gamma) = 0 \quad (4.22)$$

Equations (4.19) to (4.22) can be solved to obtain  $a$  and  $\gamma$ . A quick estimate can be found by neglecting terms involving powers of  $\gamma$  to give  $\gamma = -0.0836$  and  $a = 0.497$ . The exact solutions are  $\gamma = -0.0763$  and  $a = 0.513$ , which can be seen to be in good agreement with the simulation result of Figure 4.10.

#### 4.7.2 Effect of Bias with Odd Symmetrical Nonlinearity

The feedback system in the Simulink diagram of Figure 4.11 is considered. The odd symmetrical nonlinear characteristic has a slope of 1 for inputs between 0 and 1, 2 between 1 and 2, 1 between 2 and 4 and then zero, and is shown in Figure 4.12. The linear dynamics are  $1.1/s(s+1)^2$  and a constant value of 1.5 enters the loop at the nonlinearity output. The system is stable when the constant value is zero but its effect is to bias the nonlinearity output so that the average value of any limit cycle at the nonlinearity output has a value of 1.5. This is due to the integrator in the loop, the input of which must have an average value of zero in the steady state.

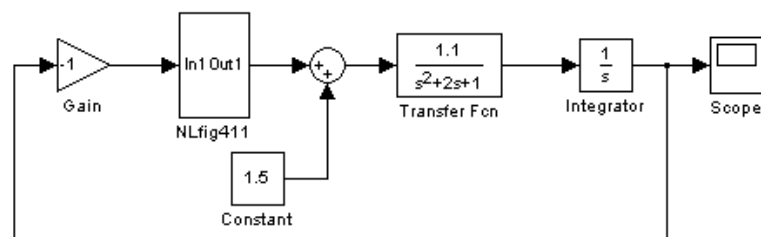


Figure 4.11 Simulink diagram of Example in section 4.7.2

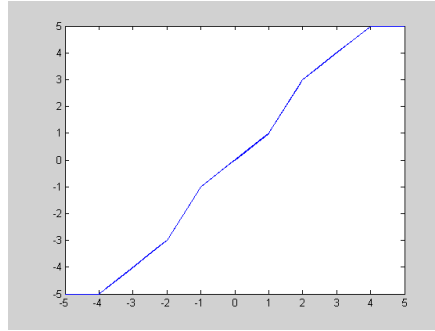
Please click the advert



Discover the truth at [www.deloitte.ca/careers](http://www.deloitte.ca/careers)

**Deloitte.**

© Deloitte & Touche LLP and affiliated entities.



**Figure 4.12** Nonlinearity in the Simulink diagram of Figure 4.11

The resulting limit cycle waveforms at the input and output of the nonlinearity, when an initial value of 2 is applied at the system output, are shown in Figure 4.13. To calculate the simulation result using SBDF analysis is not easy because of the linear segmented nonlinear characteristic. It can be simplified using the known simulation solution from which it can be seen that the amplitude of the limit cycle at the nonlinearity input is within the range -1 to 2 so that the SBDF just needs to be worked out for this range, where the characteristic is a gain of 2 in parallel with minus ideal saturation. The bias input is such that it lies on the saturated level of the saturation characteristic, thus  $\gamma > 1$ . The bias and fundamental outputs can be calculated to be

$$\text{bias} = 1 - (a/\pi)(1 - [(\gamma - 1)/a]^2)^{1/2} \text{ and } \text{fund} = (2a/\pi)[(\pi/4) - (\varphi/2) - (\sin 2\varphi/4)]$$

where  $\sin \varphi = [(\gamma - 1)/a]$ . Thus for the nonlinearity the SBDF terms are

$$N_\gamma(a, \gamma) = 2 - (1/\gamma) + (a/\pi\gamma)(1 - [(\gamma - 1)/a]^2)^{1/2} \quad (4.23)$$

and

$$N_a(a, \gamma) = 2 - (2/\pi)[(\pi/4) - (\varphi/2) - (\sin 2\varphi/4)] \quad (4.24)$$

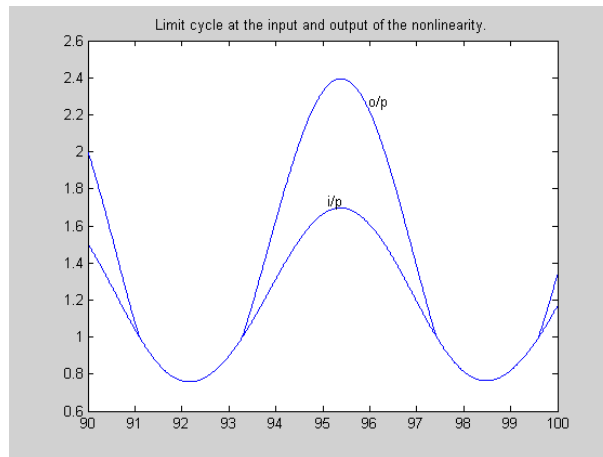
$G(j\omega)$  has a phase shift of  $180^\circ$  for  $\omega = 1$ , with a corresponding gain of 0.55, so that equation (4.16) gives

$$N_a(a, \gamma) = 1/0.55 \quad (4.25)$$

and the bias equation is

$$\gamma N_\gamma(a, \gamma) = 1.5 \quad (4.26)$$

The solution to equations (4.23) to (4.26) gives  $\gamma = 1.20$  and  $a = 0.379$  which compares reasonable well with the simulation result shown in Figure 4.13 for the nonlinearity input waveform.



**Figure 4.13** Simulation result for the example of section 4.7.2

## 4.8 Conclusions

This section has shown how limit cycles can be estimated in nonlinear systems using describing function methods. Systems where bias signals may exist in the loop have been evaluated using the SBDF. It has further been shown how the stability of any limit cycles can be assessed. Further, it has been demonstrated that when single valued nonlinear elements exist in a feedback system failure of the DF method to predict any limit cycles is due to their waveform not being 'sufficiently sinusoidal' at a nonlinearity input so that the fundamental signal undergoes a phase shift in passing through the nonlinearity which is not predicted by the DF. Many textbooks cover some basic uses of the DF usually for evaluating odd symmetrical limit cycles and the stability of limit cycles is normally commented on in terms of the Loeb criterion. Some of the more detailed topics in this chapter can be found in the books given in the bibliography.

## 4.9 Bibliography

- Atherton, DP 1975, *Nonlinear Control Engineering: Describing Function Analysis and Design*, Van Nostrand Reinhold, London.
- Atherton DP 1982, *Nonlinear Control Engineering. Student edition*, Van Nostrand Reinhold, London .
- Gelb, A & Vander Velde, WE 1968, *Multiple-Input Describing Functions and Nonlinear System Design*, McGraw Hill, New York.



# 5 The SSDF and Harmonically Forced Systems

## 5.1 Introduction

Describing functions have so far been considered for a single sinusoid and a sinusoid plus bias. To examine various phenomena which may occur in nonlinear systems it is however required to evaluate describing functions for other inputs, in particular two sinusoidal inputs. The evaluation of the sinusoidal inputs, SSDF, for any nonlinearity is considered after introduction of the topic by use of a cubic nonlinearity. A general theory is presented but the difficult algebra for specific nonlinearities makes evaluation difficult for many nonlinearities, specifically linear segmented ones, so that numerical approaches have usually to be adopted to obtain solutions. The problem is shown to be even more complicated if the sinusoidal signals are related in frequency, a situation which often needs to be considered in practice. The concept of modified nonlinearities, as a means of understanding the evaluation of the SSDF, is then covered and examples given for specific nonlinearities. Sometimes it is necessary to consider a perturbation which is related in frequency to an existing sinusoid so the evaluation of this IDF is considered next. This is followed by a practical example of the use of the SSDF to evaluate a limit cycle using a two harmonic balance.

Please click the advert

It's only an opportunity if you act on it

IKEA.SE/STUDENT

© Inter IKEA Systems B.V. 2009

The closed loop frequency response of a nonlinear system is then considered. The most usual situation is for the system to have an output with a dominant fundamental component of the same frequency as the input. This situation is examined using the DF and it is shown that over some ranges of amplitude and frequency of the input three analytical solutions rather than one to the nonlinear equations may be obtained. In practice this results in the so called ‘jump phenomena’ or ‘jump resonance’ in the frequency response. Finally it is shown how its existence can be investigated using the related frequency IDE. Various other possible behaviours that may occur, such as the excitation of oscillations and chaotic motion, are discussed based on an understanding of the SSDF and some simulation examples given. Brief mention is also made of the possible behaviours when a harmonic input is applied to an autonomous system which has a limit cycle.

Analysis of a specific system using the methods presented in this chapter is not easy unless the techniques are implemented in appropriate numerical programs, however, presentation of the theoretical background should allow the reader to understand the mechanisms that cause the aforementioned unique behaviours of nonlinear feedback systems.

## 5.2 The Cubic Nonlinearity with Two Sinusoidal Inputs

Consider the two sinusoidal inputs  $x = a \cos \theta_a$  and  $y = b \cos(\theta_b + \varphi)$ , where  $\theta_a = \omega_a t$  and  $\theta_b = \omega_b t$  into a cubic nonlinearity, then the output,  $f(t)$ , is given by

$$f(t) = [a \cos \theta_a + b \cos(\theta_b + \varphi)]^3 \quad (5.1)$$

Expanding the expression yields

$$f(t) = a^3 \cos^3 \theta_a + 3a^2 b \cos^2 \theta_a \cos(\theta_b + \varphi) + 3ab^2 \cos \theta_a \cos^2(\theta_b + \varphi) + b^3 \cos^3(\theta_b + \varphi)$$

which on expanding the powers of  $\cos \theta$  may be written

$$f(t) = (3a/4)(a^2 + 2b^2) \cos \theta_a + (3b/4)(b^2 + 2a^2) \cos(\theta_b + \varphi) + (3a^2 b/2) \cos 2\theta_a \cos(\theta_b + \varphi) \\ + (3ab^2/2) \cos \theta_a \cos 2(\theta_b + \varphi) + (a^3/4) \cos 3\theta_a + (b^3/4) \cos 3(\theta_b + \varphi)$$

and on using  $2 \cos A \cos B = \cos(A + B) + \cos(A - B)$  gives

$$f(t) = (3a/4)(a^2 + 2b^2) \cos \theta_a + (3b/4)(b^2 + 2a^2) \cos(\theta_b + \varphi) + (3a^2 b/4) \\ [\cos(2\theta_a + \theta_b + \varphi) + \cos(2\theta_a - \theta_b - \varphi)] + (3ab^2/4) [\cos(\theta_a + 2\theta_b + 2\varphi) + \cos(\theta_a - 2\theta_b - 2\varphi)] \\ + (a^3/4) \cos 3\theta_a + (b^3/4) \cos 3(\theta_b + \varphi) \quad (5.2)$$

If the frequencies  $\omega_a$  and  $\omega_b$  are unrelated then the first two terms produce the output at the same frequencies as the inputs and the SSDF has the components

$$N_a(a, b) = (3/4)(a^2 + 2b^2) \quad (5.3)$$

$$N_b(a, b) = (3/4)(b^2 + 2a^2) \quad (5.4)$$

However, if there is an integer relationship between the two frequencies then it is possible that some of the other terms in  $f(t)$  can produce outputs at the same frequency as the input. For example, if  $\omega_b = 3\omega_a$  then the second component of the third term is  $(3a^2b/4)\cos(-\theta_a - \varphi)$ , which has the same frequency as the sinusoidal input  $x$ . Further this output component is no longer in phase with the input at frequency  $\omega_a$  and the total output at this frequency is  $(3a/4)(a^2 + 2b^2)\cos\theta_a + (3a^2b/4)\cos(\theta_a + \varphi)$  so that the DF for the input  $x$ , which will be denoted  $N_a(a \mid b)$  to indicate that the signals  $a$  and  $b$  are related, is

$$N_a(a \mid b) = (3/4)(a^2 + 2b^2 + abe^{j\varphi}) \quad (5.5)$$

Thus unlike the unrelated case the above DF value depends not only on the amplitudes of the two sinusoids but also their relative phase. It is the phase shift produced through a nonlinearity to an harmonic input related to another harmonic input that results in many of the interesting phenomena in nonlinear feedback loops. Further when the nonlinearity is not a simple mathematical function, such as the cubic of the above example, it becomes very difficult to get the output by the simple trigonometric approach used above. Fortunately, however, it can be shown that for any nonlinearity  $n(x)$  the output  $f(t)$  can be written

$$f(t) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \varepsilon_j \varepsilon_k \alpha_{jk} \cos j\theta \cos k(\theta_b + \varphi) \quad (5.6)$$

where

$$\alpha_{jk} = \int_{-a}^a \int_{-b}^b n(x+y) T_j(x/a) T_k(y/b) p_a(x) p_b(y) dx dy \quad (5.7)$$

and the Neumann factor  $\varepsilon_n$  is 1 for  $n = 0$  and 2 for  $n \geq 1$ .

Comparing equation (5.6) with the result of equation (5.2) for the cubic it can be seen for the cubic that

$$\alpha_{10} = (3a/8)(a^2 + 2b^2), \alpha_{21} = 3a^2b/8, \alpha_{30} = a^3/8$$

and the values for  $\alpha_{01}, \alpha_{12}, \alpha_{03}$  can be obtained by reversing  $a$  and  $b$ . These values can be checked to result from equation (5.7) with  $n(x+y) = (x+y)^3$  and it will also be found that  $\alpha_{jk} = 0$  for  $j+k > 3$ .

Thus even though the expression of equation (5.6) can in principle be evaluated for any nonlinear characteristic it can consist of a large, and possibly infinite number of terms. Also for related input frequencies there may be a large number of these terms that produce an output at one of the input frequencies all of which in general will have a different phase.

### 5.3 Modified Nonlinearities and the SSDF

The evaluation of  $\alpha_{jk}$  from equation (5.7) can be regarded as the two stage process

$$n(a, \gamma)_j = \int_{-a}^a n(x + \gamma) T_j(x/a) p_a(x) dx \quad (5.8)$$

$$\alpha_{jk} = \int_{-b}^b n(a, \gamma)_j T_k(y/b) p_b(y) dy \quad (5.9)$$

Or, alternatively

$$n(b, \gamma)_k = \int_{-a}^a n(y + \gamma) T_k(y/b) p_b(y) dy \quad (5.10)$$

$$\alpha_{jk} = \int_{-b}^b n(b, \gamma)_k T_j(x/a) p_a(x) dx \quad (5.11)$$

The terms  $n(a, \gamma)_j$  and  $n(b, \gamma)_k$  may be regarded as generalized modified nonlinearities with the one with subscript zero, already having been introduced in section 3.7, using a notation omitting the zero subscript, which will continue to be used. Assuming, where unless mentioned to the contrary the two sinusoids are not related, then the SSDF components are  $N_a(a, b) = 2\alpha_{10}/a$  and  $N_b(a, b) = 2\alpha_{01}/b$ . An advantage of looking at the derivation in this way is that if one wishes to get the SSDF component for  $x$ , then one calculates the DF for the original nonlinearity modified by  $y$ , namely the DF of  $n(b, \gamma)$ , or alternatively one calculates  $n(a, \gamma)_1 = \int_{-a}^a n(x + y)(x/a) p_a(x) dx$ , which is  $0.5a$  times the DF for  $x$  in the presence of a bias and then averages this with respect to  $b$ , as given by equation (5.9) with  $k = 0$ .

The evaluation of modified nonlinearities for linear segmented characteristics is quite difficult and therefore it becomes even more difficult to find the SSDF, so that numerical approaches have to be used. However many nonlinear characteristics can be approximated over a given range by either a power series or Fourier series expansion, so results are derived below for a power law and harmonic nonlinearity, as well as for an ideal relay.

Please click the advert

## YOUR CHANCE TO CHANGE THE WORLD

Here at Ericsson we have a deep rooted belief that the innovations we make on a daily basis can have a profound effect on making the world a better place for people, business and society. Join us.

In Germany we are especially looking for graduates as Integration Engineers for

- Radio Access and IP Networks
- IMS and IPTV

We are looking forward to getting your application!  
To apply and for all current job openings please visit our web page: [www.ericsson.com/careers](http://www.ericsson.com/careers)

ericsson.  
com



### 5.3.1 Power Law Characteristic

Consider the nonlinearity  $n(x) = x^m$ . Then one has

$$n(a, \gamma)_k = \int_{-a}^a (x + \gamma)^m T_k(x/a) p_a(x) dx \quad (5.12)$$

which on expanding gives

$$n(a, \gamma)_k = \sum_{n=0}^m \int_{-a}^a {}_m C_n x^n \gamma^{m-n} T_k(x/a) p_a(x) dx \quad (5.13)$$

where the binomial coefficient  ${}_m C_n = m! / (m-n)! n!$ . Thus the integral can be computed for any chosen value of  $k$ , since after substituting for  $T_k(x/a)$ , the integral will involve moments of the distribution of  $p_a(x)$ . In particular for  $k = 0$ , one has

$$n(a, \gamma) = \sum_{n=0}^m \int_{-a}^a {}_m C_n x^n \gamma^{m-n} p_a(x) dx = \sum_{n=0}^m {}_m C_n \mu_n(a) \gamma^{m-n} \quad (5.14)$$

and for  $k = 1$

$$n(a, \gamma)_1 = \sum_{n=0}^m \int_{-a}^a {}_m C_n x^{n+1} \gamma^{m-n} a^{-1} p_a(x) dx = \sum_{n=0}^m {}_m C_n a^{-1} \mu_{n+1}(a) \gamma^{m-n} \quad (5.15)$$

Using equation (5.14) in equation (5.09) and multiplying by  $2/b$  to get  $N_b(b, \gamma)$  from  $\alpha_{01}$  then gives for the SSDF to the signal  $y$

$$N_b(a, b) = (2/b) \int_{-a}^b \sum_{n=0}^m {}_m C_n \mu_n(a) y^{m-n} b^{-1} y p_b(y) dy = (2/b^2) \sum_{n=0}^m {}_m C_n \mu_n(a) \mu_{m-n+1}(b) \quad (5.16)$$

Or, alternatively, equation (5.15) with  $b$  replacing  $a$  can be used in equation (5.11) and again multiplied by  $2/b$  to get  $N_b(b, \gamma)$  from  $\alpha_{01}$  to give

$$N_b(a, b) = (2/b) \int_{-a}^a \sum_{n=0}^m {}_m C_n b^{-1} \mu_{n+1}(b) x^{m-n} p_a(x) dx = (2/b^2) \sum_{n=0}^m {}_m C_n \mu_{n+1}(b) \mu_{m-n}(a) \quad (5.17)$$

Since  ${}_m C_n = {}_m C_{m-n}$  these expressions are easily shown to be identical.  $N_a(a, b)$  can be similarly found or simply obtained from  $N_b(a, b)$  by interchanging  $a$  and  $b$ .

### 5.3.2 Harmonic Nonlinearity

For the harmonic nonlinearity  $n(x) = \sin mx$

$$n(a, \gamma)_k = \int_{-a}^a \sin m(x + \gamma) T_k(x/a) p_a(x) dx$$

which on expanding  $\sin m(x + \gamma)$  and substituting  $x = a \cos \theta$  gives

$$n(a, \gamma)_k = (1/\pi) \left\{ \int_0^\pi \cos m\gamma \sin(ma \cos \theta) \cos k\theta d\theta + \int_0^\pi \sin m\gamma \cos(ma \cos \theta) \cos k\theta d\theta \right\}.$$

Substituting the expressions

$$\cos(z \cos \theta) = J_0(z) + 2 \sum_{i=1}^{\infty} (-1)^i J_{2i}(z) \cos 2i\theta \text{ and}$$

$$\sin(z \cos \theta) = -2 \sum_{i=1}^{\infty} (-1)^i J_{2i-1}(z) \cos(2i-1)\theta \text{ and integrating gives}$$

$$n(a, \gamma)_k = (-1)^k J_k(ma) \sin\{m\gamma - (k\pi/2)\} \quad (5.18)$$

In particular for  $k = 0$  the modified nonlinearity is

$$n(a, \gamma) = J_0(ma) \sin m\gamma \quad (5.19)$$

which differs only in amplitude by a factor  $J_0(ma)$  from the original nonlinearity. Also

$$n(a, \gamma)_1 = J_1(ma) \cos m\gamma \quad (5.20)$$

It is easily shown, again using the above expansions, that

$$N_b(a, b) = (2/b) J_0(ma) J_1(mb) \quad (5.21)$$

### 5.3.3 Ideal Relay

The two previous characteristics considered were continuous functions of the input variable, whereas this is not the case for many approximations to nonlinearities found in control systems, such as the relay or ideal saturation. These create difficulties for obtaining general analytical results for the SSDF because the modified nonlinearity will possess discontinuities dependent on the relative input magnitudes of the signal and the bias. This is illustrated for the ideal relay with output level  $h$  for which

$$n(a, \gamma)_k = \int_{-a}^{-\gamma} -h T_k(x/a) p_a(x) dx + \int_{-\gamma}^a h T_k(x/a) p_a(x) dx$$

which gives for  $k \geq 1$

$$n(a, \gamma)_k = (2h/k\pi) D_k(-\gamma/a) \text{ for } |\gamma| \leq a \quad (5.22)$$

$$n(a, \gamma)_k = 0 \text{ for } |\gamma| > a$$

and for  $k = 0$

$$n(a, \gamma) = (2h/\pi) \sin^{-1}(\gamma/a) \text{ for } |\gamma| \leq a \quad (5.23)$$

$$n(a, \gamma) = h \operatorname{sgn} \gamma \text{ for } |\gamma| > a$$

If it is assumed that  $b \leq a$  then to find  $N_b(a, b)$  only  $n(a, y)$  is required for  $|y| \leq a$  and

$$N_b(a, b) = (2/b) \int_{-b}^b n(a, y)(y/b)p_b(y) dy = (4h/\pi^2 b^2) \int_{-b}^b y \sin^{-1}(y/a)(b^2 - y^2)^{-1/2} dy$$

Integrating by parts and using the fact that  $(b^2 - y^2)^{1/2}$  is zero at both limits gives

$$N_b(a, b) = (4h/\pi^2 b^2) \int_{-b}^b \frac{(b^2 - y^2)^{1/2}}{(a^2 - y^2)^{1/2}} dy = (8h/\pi^2 k^2 a) \{E(k) - (1 - k^2)K(k)\} \quad (5.24)$$

where  $k = b/a < 1$  and  $E(k)$  and  $K(k)$  are elliptic integrals. Similarly, to find  $N_a(a, b)$  only one integral is required for  $b \leq a$  if  $n(a, y)_1$  is used and the solution is

$$N_a(a, b) = (4h/a\pi^2) \int_{-b}^b \frac{[1 - (y/a)^2]}{(b^2 - y^2)^{1/2}} dy = (8h/a\pi^2)E(k) \quad (5.25)$$

Thus even for this very simple break point type nonlinearity the SSDF expressions are rather complicated.

## 5.4 The IDF for Related Signals

The IDF, the gain to an unrelated signal in the presence of a known other one, was considered in section 3.7 for the known signal being a sinusoid. Also the perturbation equation of (4.5) to assess stability was derived assuming the perturbation was unrelated to the existing signal. Here the IDF is investigated when the small signal is assumed to be related to the sinusoid. Consider a nonlinearity with inputs  $x(t) = a \cos \theta_a$  and a small signal  $\delta(t)$ . The nonlinearity output, making use of a Taylor series expansion, as used to derive equation (4.5), may be written as  $n\{x(t) + \delta(t)\} \approx n(x(t)) + \delta(t)n'[x(t)]$  where  $n'(x) = dn(x)/dx$ .

The incremental output,  $\delta y(t)$  due to  $\delta(t)$  is thus  $\delta(t)n'[x(t)]$  the output of the nonlinearity  $n'(x)$  with input  $x(t)$ . This output may be written for any nonlinearity in the Fourier series.

$$n'\{x(t)\} = \sum_{k=0}^{\infty} c_k \cos(k\theta_a + \varphi_k)$$

where  $x(t)$  has a frequency of  $\omega_a = \theta_a t$ . If  $\delta(t)$  is the small sinusoid  $\delta \cos(\theta_b + \phi)$ , with  $\omega_b = \theta_b t$ , then

$$\delta y(t) = \delta \sum_{k=0}^{\infty} c_k \cos(k\theta_a + \varphi_k) \cos(\theta_b + \phi) \quad \text{which can be written}$$

$$\delta y(t) = \delta \sum_{k=0}^{\infty} (c_k/2) \{ \cos(k\theta_a + \varphi_k + \theta_b + \phi) + \cos(k\theta_a + \varphi_k - \theta_b - \phi) \}$$

The only output at  $\omega_b$  if  $\omega_b$  and  $\omega_a$  are not related is  $\delta c_0 \cos(\theta_b + \phi)$  but if  $\theta_a/\theta_b = \omega_a/\omega_b = m/n$  then the output at frequency  $\omega_b$  is

$$\delta \{ c_0 \cos(\theta_b + \phi) + (c_k/2) \cos(\theta_b + \varphi_k - \phi) \mid_{k=2n/m} \} \quad (5.26)$$

which can be written



$$\delta c_0 \cos(\theta_b + \phi) + (\delta c_k/2) \{ \cos(\varphi_k - 2\phi) \cos(\theta_b + \phi) - \sin(\varphi_k - 2\phi) \sin(\theta_b + \phi) \} \mid_{k=2n/m}$$

Thus, dividing by  $\delta$ , gives the IDF for situations where the frequencies are related such that if an integer  $k = 2n/m$  exists, then it is given by

$$N_{i\delta}(a) = c_0 + (c_k/2) e^{j(\varphi_k - 2\phi)} \mid_{k=2n/m} \quad (5.27)$$

where the subscript  $\delta$  has been used to distinguish it from the unrelated IDF  $N_{iy}(a)$ . Since in general the coefficients  $c_k$  normally decrease with increasing  $k$ , the most important ones are for low values of  $k$ . Those for  $k = 1$  and  $2$ , in particular, will be of some interest later. From equation (5.27) it can be seen that  $N_{i\delta}(a)$  is complex and its locus is a circle centred at  $c_0$  with radius  $c_k/2$ . If one wishes to plot  $C_{i\delta}(a) = -1/N_{i\delta}(a)$  on a Nyquist plot then it can be shown that  $C_{i\delta}(a)$  is also a circle with centre at  $-c_0 / \{c_0^2 + (c_k/2)^2\}$  and radius  $(c_k/2) \{c_0^2 + (c_k/2)^2\}$ .

Also for a single valued nonlinearity with  $k = 2$ , which is the situation for a perturbation of the same frequency,  $\varphi_k = 0$  and

$$c_2 = 2 \int_{-a}^a n'(x) [(2x^2/a^2) - 1] p(x) dx = 2 [N_{iy}(a) - N(a)]$$


so that

$$N_{i\delta}(a) = N(a) + (a/2) N'(a) (1 + e^{-j2\phi}) \quad (5.28)$$


This is the equation of a circle with centre at  $(N_{iy}(a), 0)$  and radius  $(a/2) N'(a)$ .

Please click the advert

SIMPLY CLEVER



We will turn your CV into  
an opportunity of a lifetime



Do you like cars? Would you like to be a part of a successful brand?  
We will appreciate and reward both your enthusiasm and talent.  
Send us your CV. You will be surprised where it can take you.

Send us your CV on  
[www.employerforlife.com](http://www.employerforlife.com)



## 5.5 More Accurate Determination of Limit Cycles

The classical approach is to balance more harmonics not just the fundamental. Although straightforward in principle it is not easy analytically due to the difficulty of computing the output from a nonlinear element subject to two or more related harmonic inputs. To illustrate the procedure a feedback loop with the simple polynomial nonlinearity  $n(x) = x - (x^3/6)$  and a loop transfer function  $G(s) = 2(1 - s)/s(1 + s)$  is considered.

The nonlinearity input is taken as

$$x = a \cos \omega t + b \cos(3\omega t + \varphi) \quad (5.29)$$

Computing the output from the nonlinearity at frequencies  $\omega$  and  $3\omega$ , gives the terms

$$\left[ a - (a/8)(a^2 + 2b^2 + ab \cos \varphi) \right] \cos \omega t + (a^2 b/8) \sin \varphi \sin \omega t \quad (5.30)$$

and

$$\left[ b - (b/8)(b^2 + 2a^2) - (a^3/24) \cos \varphi \right] \cos(3\omega t + \varphi) - (a^3/24) \sin \varphi \sin(3\omega t + \varphi) \quad (5.31)$$

Thus the magnitude and phase components for the SSDF are

$$M_a(a|b) = \left\{ \left[ 1 - (1/8)(a^2 + 2b^2 + ab \cos \varphi) \right]^2 + \left[ (1/8)ab \sin \varphi \right]^2 \right\}^{1/2} \quad (5.32)$$

$$M_b(a|b) = (1/b) \left\{ \left[ b - (b/8)(b^2 + 2a^2) + (a^3/24) \cos \varphi \right]^2 + \left[ (a^3/24) \sin \varphi \right]^2 \right\}^{1/2} \quad (5.33)$$

$$\phi_a(a|b) = \tan^{-1} \left[ -ab \sin \varphi / (8 - a^2 - 2b^2 - ab \cos \varphi) \right] \quad (5.34)$$

$$\phi_b(a|b) = \tan^{-1} \left[ a^3 \sin \varphi / (24b - 3b^3 - 6a^2 b - a^3 \cos \varphi) \right] \quad (5.35)$$

If the gain and phase of the transfer function are denoted by  $g$  and  $\theta$  then the harmonic balance loop equations for the two frequencies are

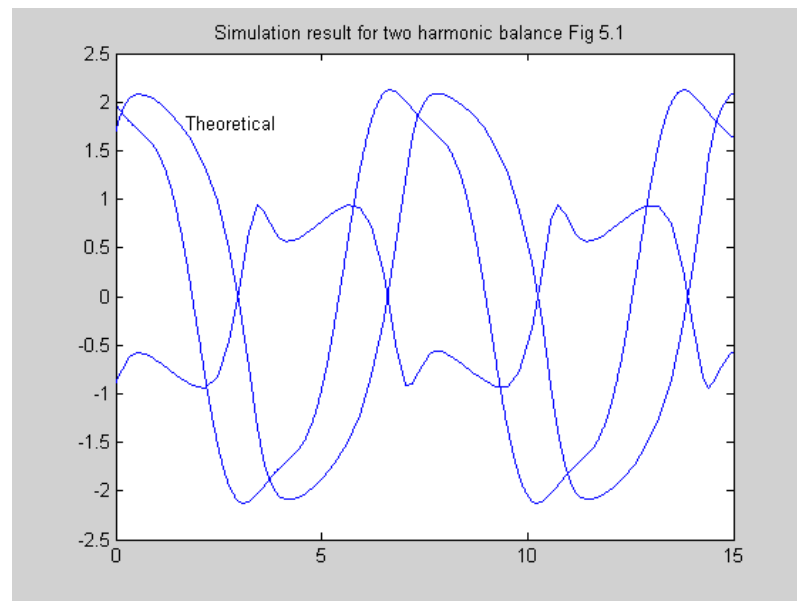
$$M_a(a|b)g(\omega) = 1 \quad (5.36)$$

$$M_b(a|b)g(3\omega) = 1 \quad (5.37)$$

$$\phi_a(a|b) + \theta(\omega) = -180^\circ \quad (5.38)$$

$$\phi_b(a|b) + \theta(3\omega) = (2j + 1)180^\circ \quad (5.39)$$

An estimate for the solution can be obtained by using the DF solution of  $a = 2$  and  $\omega = 1$ , together with equations (5.37) and (5.39) with (5.33) and (5.35) substituted. This gives  $b = 0.222$  and  $\varphi = -53^\circ$  or  $127^\circ$ . Solution of the four nonlinear equations (5.36) to (5.39) with the values from equations (5.32) to (5.35) substituted gives two solutions of  $a = 2.19$ ,  $b = 0.298$ ,  $\varphi = 138^\circ$ ,  $\omega = 0.882$  and  $a = 1.88$ ,  $b = 0.166$ ,  $\varphi = -52^\circ$ ,  $\omega = 1.06$ . The former solution is that of the stable limit cycle shown in the simulation result of Figure 5.1. Also shown are the simulation results for the output of the nonlinearity and the calculated result for the stable limit cycle marked theoretical. Note the phase origin is arbitrary so the waveform has been plotted just to show the good waveform agreement.



**Figure 5.1** Comparison of theoretical and simulation limit cycle results.

It is relatively easy to write numerical routines to compute more precisely any limit cycles which may exist in a feedback loop. References 5.1 to 5.3 give details of some papers which discuss possible approaches using programs which make use of the fast Fourier transform and interactive graphics. Because a large number of nonlinear algebraic equations need to be solved an initial guess for the limit cycle is required. This can be obtained using the DF, if a predicted solution exists, or if not by assuming a solution using say an amplitude and frequency near to an intersection of the Nyquist plot and the  $C(a)$  locus. The assumed limit cycle solution,  $x_1(t)$ , is then used as input to the nonlinearity with the loop assumed open and the signal fed back to this point  $x_2(t)$  calculated. This is done by discretising the signal, finding the corresponding output from the nonlinearity, evaluating its Fourier series, passing these components through  $G(s)$  and recombining them to produce the nonlinearity input. Graphs of  $x_1(t)$  and  $x_2(t)$  are then plotted for comparative purposes. Further iteration can be done to see if the waveforms converge, typically they may become quite similar but because the assumed frequency is incorrect appear to have a phase difference.

The frequency can then be changed to see if convergence occurs. Alternatively at any point in the process a harmonic balance may be computed using a limited number of harmonics in  $x_1(t)$  by taking it equal to the last value of  $x_2(t)$ . Using these techniques examples are given in the papers referenced of deriving limit cycles in a large number of feedback systems some with multiple nonlinear and linear elements. Cases are considered when the DF method does and does not predict a limit cycle. An interesting feature of the approach is that because it is numerical it can obtain unstable limit cycle solutions which cannot be found by simulation. One example which was considered is the feedback loop originally investigated by Fitts with a cubic nonlinearity and linear dynamics

$$G(s) = 10s^2 / (s^4 + 0.04s^3 + 2.0206s^2 + 0.0404s + 0.9803). \quad (5.40)$$

This transfer function, which is not typical of one found for a control system with the double derivative in the numerator, has two pairs of lightly damped poles near to frequencies of 0.9 and 1.1 rads/s. and its Nyquist plot is shown in Figure 5.2. The feature of this Nyquist plot, which causes the DF method to fail, is that it has related frequencies on different sides of the  $C(a)$  plot, which for a cubic lies on the negative real axis. The Nyquist plot starts from the origin and loops in the positive part of the plane to the real axis which it reaches at a frequency of 1.005 before looping in the negative half plane to finish at the origin.

Please click the advert

# With us you can shape the future. Every single day.

For more information go to:  
[www.eon-career.com](http://www.eon-career.com)

Your energy shapes the future.



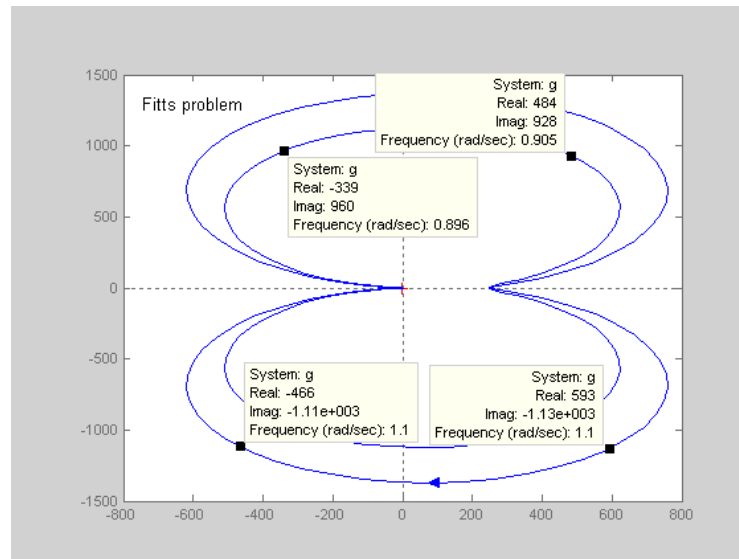


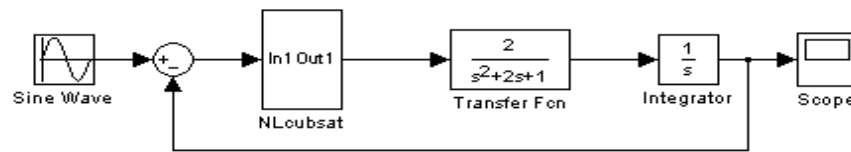
Figure 5.2 Nyquist plot for transfer function of equation (5.40)

In chapter 6 it is shown how limit cycles can be calculated exactly in feedback loops with a relay nonlinear element. Any nonlinear characteristic can be approximated by a quantization characteristic, the more quantization levels used the more accurate the approximation. Since a quantization characteristic can be obtained by relays in parallel the method for finding limit cycles in relay systems can be used for any nonlinearity using a quantized approximation. The problem with the method is that the number of nonlinear equations which have to be solved increases with the number of quantization levels used. However, by slowly increasing the number of quantization levels good initial estimates can be obtained for the nonlinear equations and good results are reported in reference 5.4 using this approach.

## 5.6 Closed Loop Frequency Response

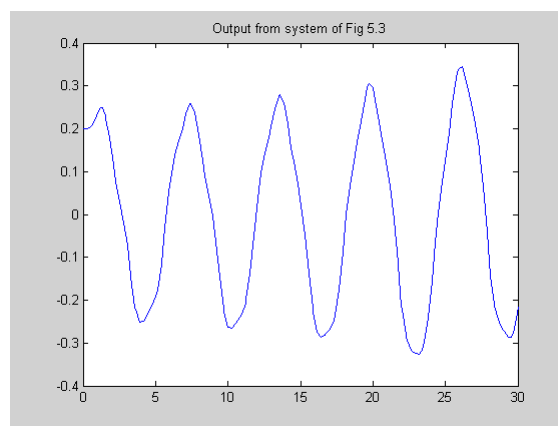
When a sinusoidal input is applied to a feedback loop which is linear then provided it is stable the output will also be a sinusoid with amplitude and phase dependent on the input frequency. Further if the input amplitude is doubled then this just doubles the output amplitude. Obviously if the system is nonlinear then the frequency response will be amplitude dependent but an obvious question is will the output be roughly sinusoidal and at the same frequency as the input? This aspect will be examined in the next section, 5.7, by assuming that it will be the case but before doing so it is appropriate to comment on some other possibilities suggested by the previously derived DF results.

Results for the SSDF have shown that both the gain and phase shift which a sinusoid undergoes in passing through a nonlinear element can be affected by another signal. If the other signal is unrelated to an existing one then only the gain to the existing one will be affected. For a limit cycle to build up in a feedback loop then the loop must be unstable to a small signal, clearly a situation which might arise when an external signal causes the gain through a nonlinear element to a small signal to change appropriately. Figure 5.3 shows a simple feedback loop with a nonlinear element  $n(x) = x^3$  for  $|x| \leq 1$ , 1 for  $|x| > 1$  and -1 for  $|x| < -1$ , dynamics of  $2/s(s+1)^2$  and a sinusoidal input of amplitude 1 at a frequency of 4.1 rad/s.



**Figure 5.3** Simple feedback loop with cubsat nonlinearity

Although the system is stable with no input, as indicated by the DF method, or as can be ascertained by simulation, when subject to the above sinusoidal input the gain through the nonlinearity is increased sufficiently to cause a limit cycle to occur at an approximate frequency of 1 rad/s, the  $180^\circ$  phase shift frequency of  $G(s)$ . The output of the system showing the dominant limit cycle frequency, which is unrelated to that of the input, is shown in Figure 5.4



**Figure 5.4** Output of System of Figure 5.3

Also the value of  $N_{i0}(a)$  for  $k = 1$ , which is achieved with  $2n = m$ , can have a phase shift. This indicates the possibility of a limit cycle occurring which is at one half of the frequency of the applied input, namely a second subharmonic. Clearly also other subharmonics might exist if they can somehow be excited since a phase change is possible for finite amplitudes of other related frequencies, such as the ratio 3:1.

A further behaviour that is possible is for the output to have a chaotic motion, that is a waveform which is deterministic, in the sense that if the experiment is repeated the same waveform will result, but it cannot be described analytically. This is demonstrated for the system of Figure 5.5 with an unusual nonlinearity  $n(x) = x - (x^3/6)$ , which changes from a positive to a negative slope. The loop dynamics are  $2/(1 + s)^3$  and the input frequency of the sinusoid is 1 rad/s. Figures 5.6 to 5.8 show the output response as the amplitude of the input sinusoid is increased. In Figure 5.6 for an input amplitude of 1.1 the output, unlike for smaller input amplitudes, is noticeably distorted from a sinusoid.

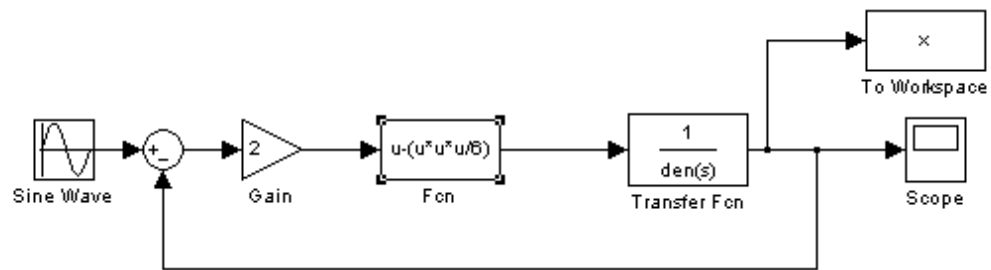


Figure 5.5 System which demonstrates chaotic motion.

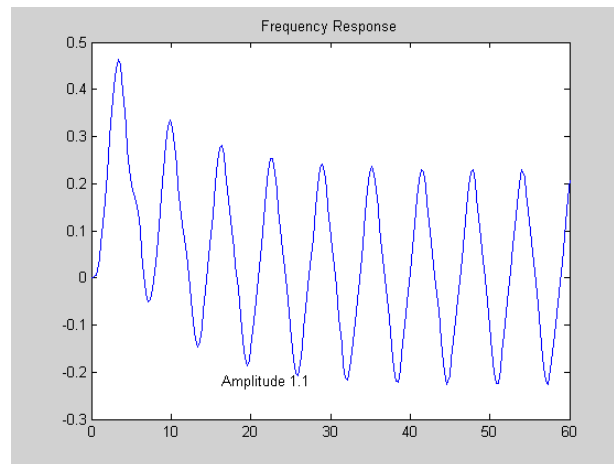


Figure 5.6 output for input amplitude of 1.1.

Please click the advert

**Nido**

**Luxurious accommodation**

**Central zone 1 & 2 locations**

**Meet hundreds of international students**

**BOOK NOW and get a £100 voucher from voucherexpress**

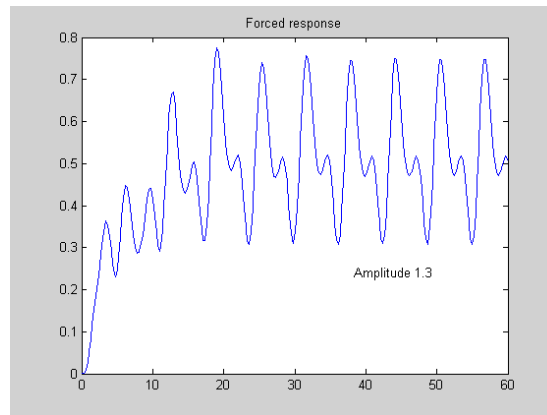
**Nido Student Living - London**

**Visit [www.NidoStudentLiving.com/Bookboon](http://www.NidoStudentLiving.com/Bookboon) for more info.**

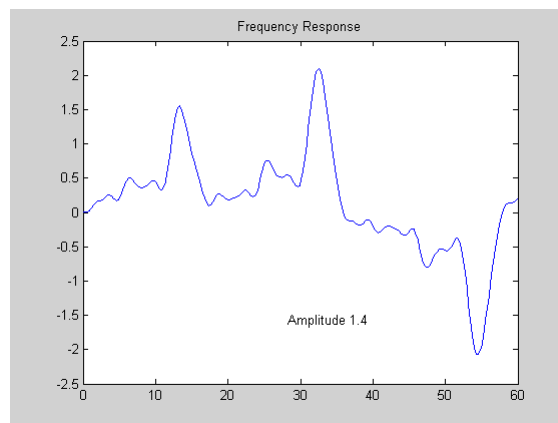
**+44 (0)20 3102 1060**



With an increase in amplitude to 1.3, shown in Figure 5.7, the output has reached a bias level and shows considerable third harmonic distortion. Finally in Figure 5.8 for an input amplitude of 1.4 the output shows a chaotic motion and for an input amplitude of around 1.5 the response goes unbounded.



**Figure 5.7** Output for input amplitude of 1.3



**Figure 5.8** Chaotic motion for an input amplitude of 1.4.

Obviously another possibility is that one may wish to apply a sinusoidal input to a system that already has a limit cycle, a situation which once more means two sinusoidal inputs to a nonlinearity need to be considered. Again based on the SSDF three possibilities can be expected namely (i) the limit cycle and forcing signal will coexist, (ii) the forcing signal may 'quench' the limit cycle or (iii) the forcing signal and limit cycle may 'lock' to related frequencies, known as synchronization.



Clearly to investigate theoretically any of these varied behaviours which may exist in a nonlinear system for a sinusoidal input one has to apply the appropriate analysis to see if such a behaviour might exist. Secondly one has then to show that the behaviour is stable. Affirmative solutions to these two questions then means, assuming the approximate analysis is sufficiently accurate, that the behaviour will probably exist. However the analysis is based on steady state assumptions and whether a specific steady state is reached is a dynamic problem which is difficult to solve. Like singular points more than one behaviour may be reached dependent on the initial conditions. Luckily most of the above behaviours are rare in a feedback control system since to exist they require features such as a lightly damped feedback loop, or ‘unusual’ nonlinearities, for example possessing both positive and negative slopes. In the next section the ‘normal’ frequency response is examined such as that found for the system of Figure 5.5 for input amplitudes less than unity.

## 5.7 Jump Resonance

When the closed loop system of Figure 3.1 has a sinusoidal input  $r(t) = R \sin(\omega t + \beta)$ , it is possible to evaluate the closed loop frequency response using the DF. If the feedback loop has no limit cycle when  $r(t) = 0$  and, in addition, the sinusoidal input  $r(t)$  does not induce a limit cycle, then, provided that  $G_c(s)G_1(s)$  gives good filtering,  $x(t)$ , the nonlinearity input may be taken equal to the sinusoid  $a \sin \omega t$ . Thus with the nonlinear element replaced by its DF, equations can be obtained for the loop operation. For example balancing the components of frequency  $\omega$  around the loop gives

$$Rg_c \sin(\omega t + \beta + \theta_c) - aM(a)g_1g_c \sin(\omega t + \phi(a) + \theta_1 + \theta_c) = a \sin \omega t \quad (5.41)$$

where  $G_c(j\omega) = g_c e^{j\theta_c}$  and  $G_1(j\omega) = g_1 e^{j\theta_1}$ . Then, in principle, equation (5.41) which can be written as two nonlinear algebraic equations, by resolving in the directions of  $\sin \omega t$  and  $\cos \omega t$ , can be solved for the two unknowns  $a$  and  $\beta$  for the given sinusoidal input with magnitude,  $R$ , and frequency,  $\omega$ . The fundamental output signal can then be found from

$$c(t) = aM(a)g_1 \sin(\omega t + \phi(a) + \theta_1) \quad (5.42)$$

Various graphical procedures [5.5-5.7] have been proposed for solving the two nonlinear algebraic equations resulting from equation (5.41) or for the loop balance results written in a slightly different form. If the system is lightly damped, the nonlinear equations may have more than one solution, indicating that the frequency response of the system may have a jump resonance. This phenomenon of a nonlinear system has been studied by many authors, both theoretically and for possible practical applications. The major advantage of a graphical approach to obtain the solution, rather than a computational approach to solving the nonlinear algebraic equations of (5.41), is the understanding it provides of the jump resonance phenomenon. One approach is briefly outlined which is applicable for a single valued nonlinearity, that is for  $\phi(a) = 0$ . From equation (5.41), on separating the  $\sin \omega t$  and  $\cos \omega t$  terms, one obtains

$$Rg_c \cos(\beta + \theta_c) - aM(a)g_1g_c \cos(\theta_1 + \theta_c) = a \quad (5.43)$$

and

$$Rg_c \sin(\beta + \theta_c) = aM(a)g_1g_c \sin(\theta_1 + \theta_c) \quad (5.44)$$

Denoting the normalized input magnitude to the nonlinearity  $a/R$  by  $x$ , the normalized output  $aM(a)/R$  by  $y$  and eliminating  $\beta$  gives

$$x = g_c \left[ 1 - y^2 g_1^2 \sin^2(\theta_1 + \theta_c) \right]^{1/2} - y g_1 g_c \cos(\theta_1 + \theta_c) \quad (5.45)$$

This is the linear frequency dependent relationship between the nonlinearity input and output which needs to be solved together with the DF relationship

$$y = xM(xR) \quad (5.46)$$

which needs to be plotted on the same x-y plot as equation (5.44) for the given value of  $R$  to get possible solutions. In cases where jump resonance occurs there are typically three solutions rather than one solution [5.8-5.10]. In this case it is possible using the stability criterion of the next section to show that two solutions will be stable and one unstable. An interesting phenomenon called island-type jump resonance is described in reference 5.11. Here one stable solution is isolated, so that it may only be reached from the application of suitable initial conditions, and therefore may not be seen in a standard frequency response test.

Please click the advert

I joined MITAS because  
I wanted **real responsibility**

The Graduate Programme  
for Engineers and Geoscientists  
**Maersk.com/Mitas**



**Month 16**

I was a construction  
supervisor in  
the North Sea  
advising and  
helping foremen  
solve problems

Real work  
International opportunities  
Three work placements





### 5.7.1 Jump resonance region- use of the IDF

Determining whether a specific system can have a jump resonance in its frequency response can be quite laborious using the method of the previous section. Based on a phasor diagram for the DF linearised closed loop system it can be argued that for a constant frequency input if its magnitude  $R$  is increased then  $a$  should also increase. Thus Hatanaka [5.12] obtained the criterion that for a system to possess jump resonance

$\left(\frac{\partial R}{\partial a}\right)_{\omega_{const}} < 0$  and the jump resonance points will be given by  $\left(\frac{\partial R}{\partial a}\right)_{\omega_{const}} = 0$ . Further results have been found using these facts but all are consistent with the fact that jump resonance will occur if a solution exists for the IDF characteristic equation

$$1 + N_{i\delta}(a)G(j\omega) = 0 \quad (5.46)$$

where the IDF is for a perturbation of the same frequency as the input as given in equation (5.28).

For a specific solution amplitude  $a$  this equation is easily checked but of course this requires the solution for  $a$  to be available. By plotting  $C_{i\delta}(a) = -1/N_{i\delta}(a)$ , which have been shown to be circles, for all possible values of  $a$  then no intersection of  $G(j\omega)$  with these loci ensures that jump resonance will not exist. From equation (5.28) the maximum phase shift produced for a given  $a$  is  $\alpha$  where

$$\sin \alpha = \frac{|(a/2)N'(a)|}{N(a) + (a/2)N'(a)} \quad (5.47)$$

If  $\alpha_m$  is the maximum value of  $\alpha$  for all  $a$  then no jump resonance will exist if  $G(j\omega)$  does not enter the sector on the Nyquist plane bounded by the radial lines  $180 \pm \alpha_m$ . Interestingly for a power law nonlinearity  $n(x) = c|x|^\mu \text{sgn}(x)$  it can be shown that

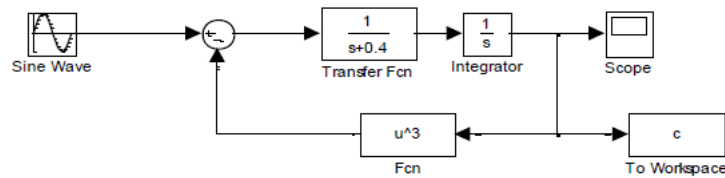
$$\alpha = \tan^{-1}[(1 - \mu)/2\mu^{1/2}] \quad (5.48)$$

which is independent of  $a$  so that  $\alpha_m = \alpha$  for all  $a$ . In particular for a cubic  $\mu = 3$  and  $\alpha = 30^\circ$ .

### 5.7.2 Some Examples of Jump Resonance

As stated above the DF analysis presented in the previous sections is based on steady state conditions. Whether a specific steady state is reachable is a dynamic problem. The existence of jump resonance behaviour was reported in the 50's and 60's from some studies, amongst others, of nonlinear electrical circuits and analogue computer simulations. The experiments were usually done using sinusoidal signal generators with the amplitude and frequency controlled manually. Typically the amplitude for the sinusoidal signal generator was set at a fixed value and the frequency varied slowly by hand turning of the frequency setting dial, first in an increasing direction to a certain value and then decreasing back to the starting frequency. Clearly there was no careful control of the dynamics of the procedure but two different levels of output were found over the 'jump range', one for increasing and the other for decreasing the frequency. Digital simulation allows more precise control of the dynamics of an experiment so for investigating the jump phenomenon questions arise as to how the sweep should be generated and how fast it should be to 'claim' the steady state has been reached.

I am indebted to H.K.Wong of the University of Warwick for the majority of the results in this section. To start the study a simple feedback loop having a transfer function  $G(s) = 1/(s + c)$  in the forward path and a cubic nonlinearity in the feedback path is considered. A Simulink diagram is shown in Figure 5.9 with  $c = 0.4$ .



**Figure 5.9** Simulink Block Diagram for Simulation Study

Denoting the input to the NL by  $a \cos \omega t$ , which is the system output, and the input to the system by  $R \cos(\omega t + \beta)$ , the loop balance equation is

$$Rg \cos(\omega t + \beta + \theta) - aM(a)g \cos(\omega t + \theta) = a \cos \omega t \quad (5.49)$$

giving

$$Rg \cos(\beta + \theta) - aM(a)g \cos \theta = a \quad (5.50)$$

and

$$-R \sin(\beta + \theta) + aM(a) \sin \theta = 0. \quad (5.51)$$

where  $G(j\omega) = g(\omega)e^{j\theta(\omega)}$  and the dependence of  $g$  and  $\theta$  on  $\omega$  has been omitted.

Denoting  $aM(a)/R$  as  $y$  and  $a/R$  as  $x$  and eliminating  $\beta$  from the above equations gives

$$y^2 + (x^2/g^2) - (2xy \cos \theta/g) = 1 \quad (5.52)$$

To show graphically the solutions this  $x$ - $y$  relationship for the frequency response of  $G$  is plotted for different values of  $\omega$  and the  $x$ - $y$  relationship for the cubic is then superimposed on it. The results are shown in Figure 5.10 for a range of  $\omega$  together with the cubic relationship for  $R=1$ , which is simply  $y = x * M(x)$  and is  $y = 0.75x^3$  for the cubic. By plotting the relationship in this way the possible solutions are easily seen for a fixed  $R$  and if  $R$  is changed only the single curve of the nonlinear relationship needs to be changed.

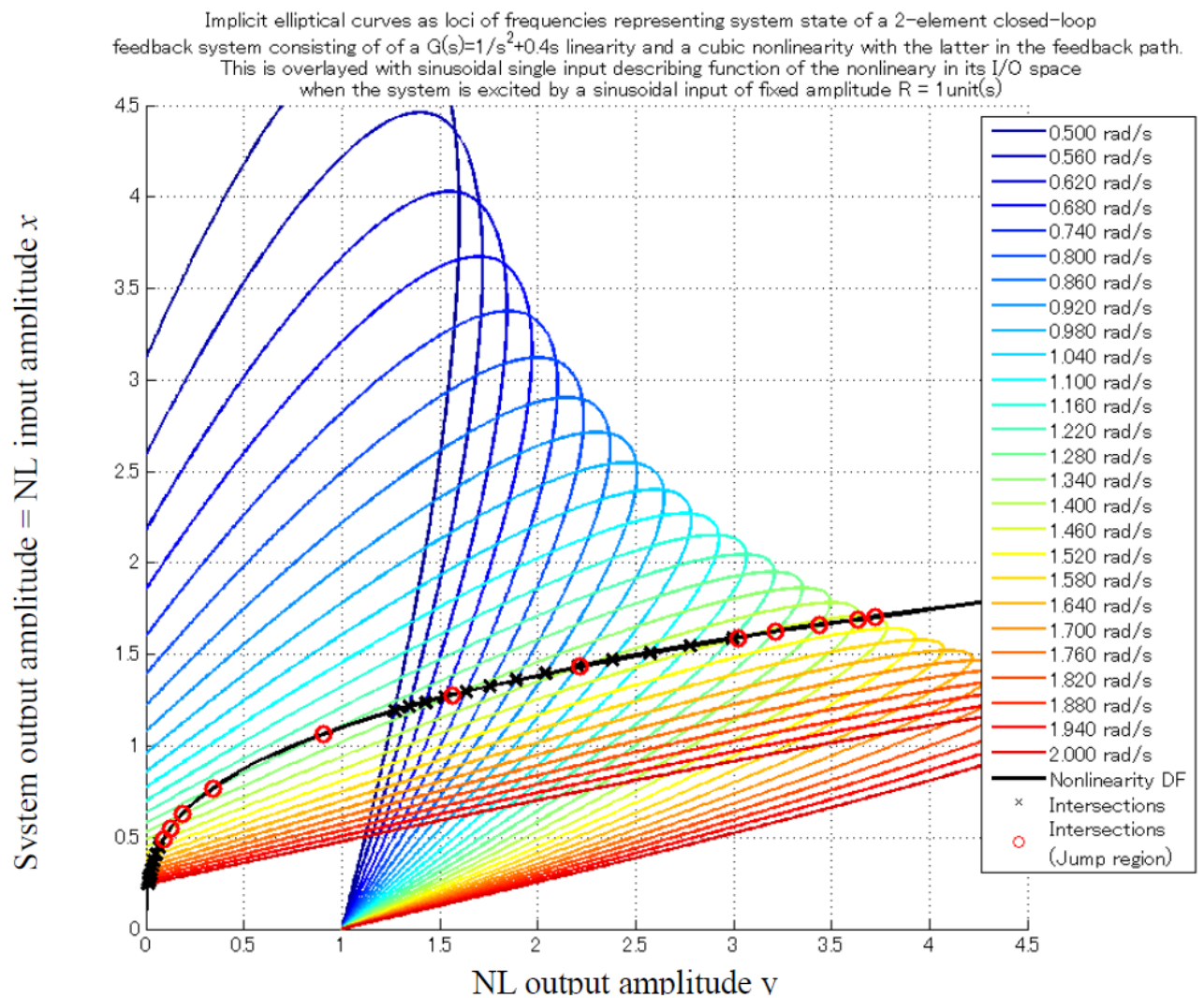


Figure 5.10 Ellipses for  $G(j)$  with  $c = 0.4$  and cubic nonlinearity for system of Figure 5.9

The intersections between each ellipse for a particular frequency with the nonlinearity input-output ( $x$ - $y$ ) plot mark the amplitude solutions for the fundamental in this  $x$ - $y$  space. The coordinates of the intersections can then be used to compute the magnitude or phase response of the system. Jump phenomena may only occur in the frequency range where there is more than one intersection of the nonlinear characteristic with an ellipse for that frequency

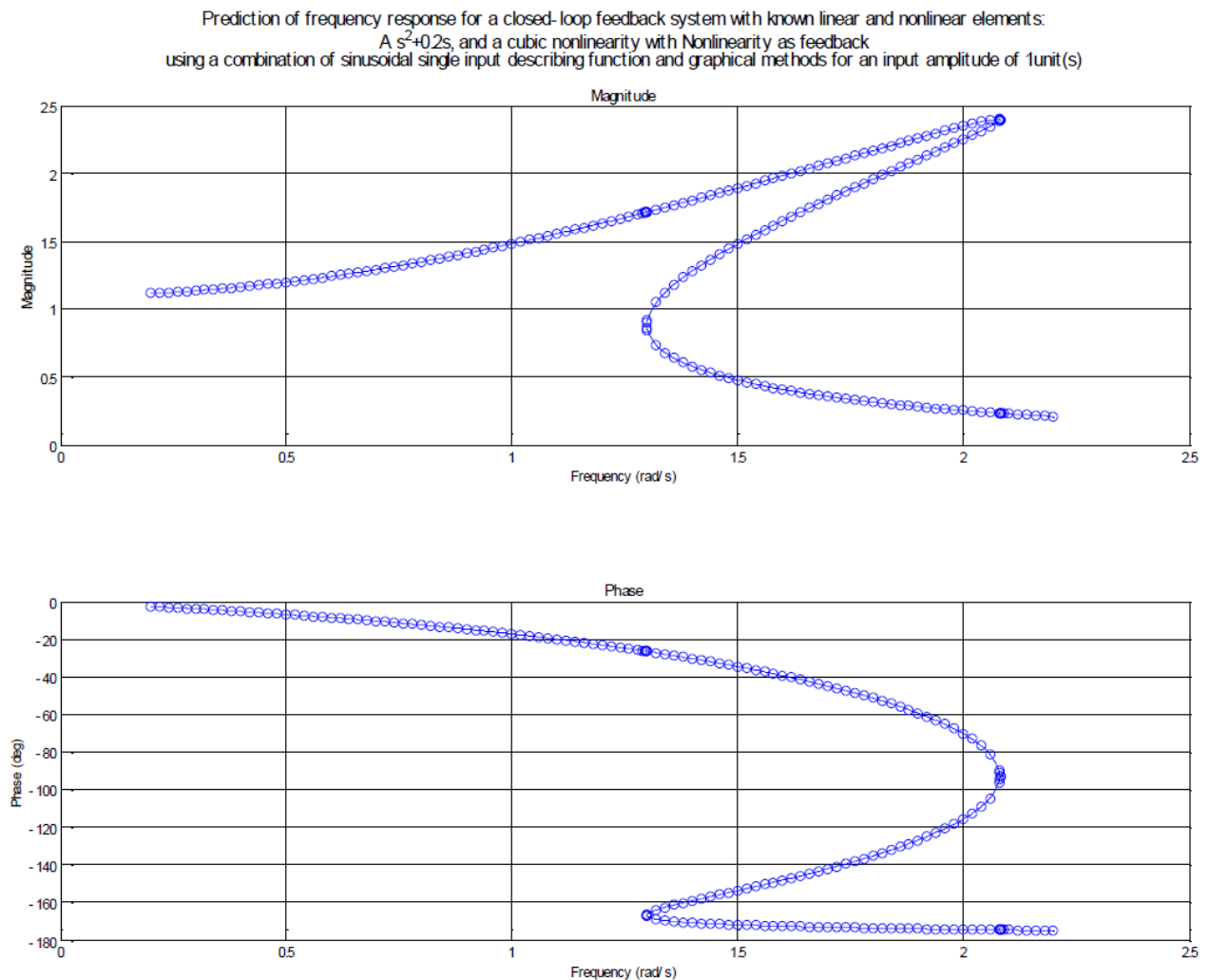
Substituting  $M(a) = 3a^2/4$  for the cubic in equations (5.50) and (5.51) gives

$$Rg \cos(\beta + \theta) - 0.75a^3 g \cos \theta = a \quad (5.53)$$

and

$$-R \sin(\beta + \theta) + 0.75a^3 \sin \theta = 0 \quad (5.54)$$

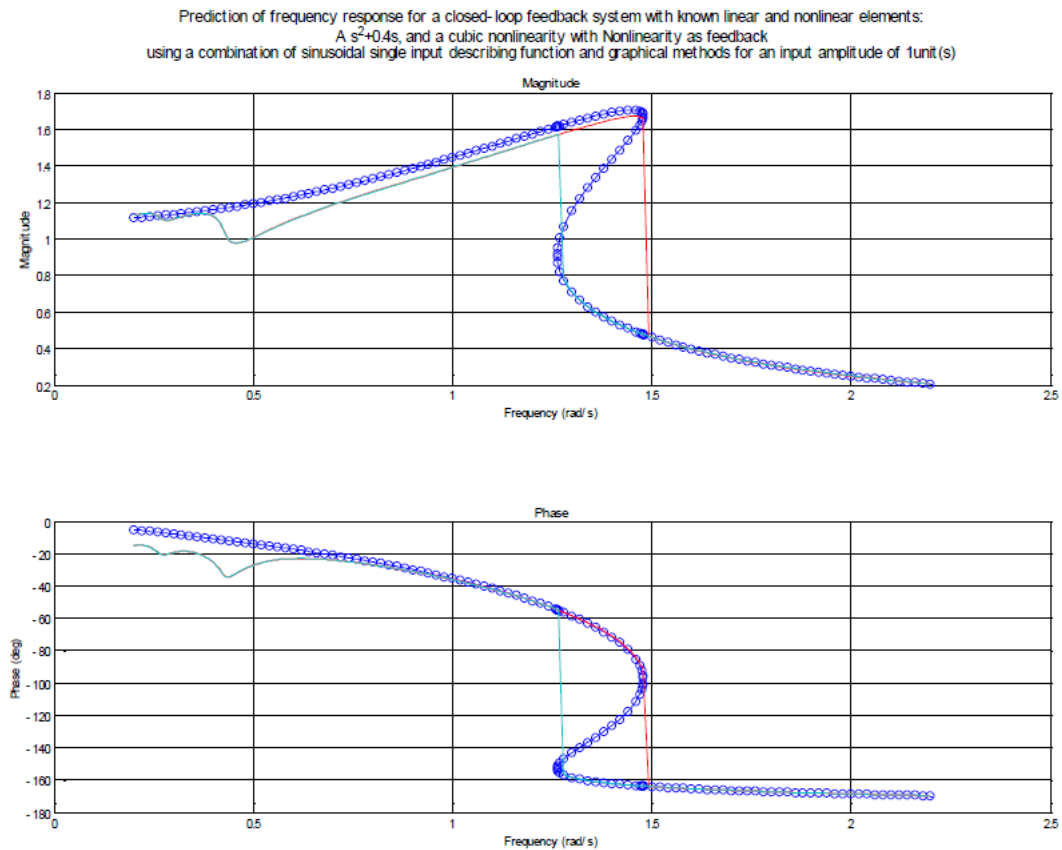
These equations can be solved numerically for  $a$  and  $\beta$  for given  $R$  and  $\omega$  avoiding the graphical approach. However, initial guesses for the solution values have to be given to a nonlinear algebraic equation solver, which is greatly facilitated by knowledge from the graphical approach. From software written using the graphical approach [5.13] the theoretical frequency response predictions are given in Figure 5.11 for  $c = 0.2$  in the transfer function  $G$  and input  $R = 1$ .



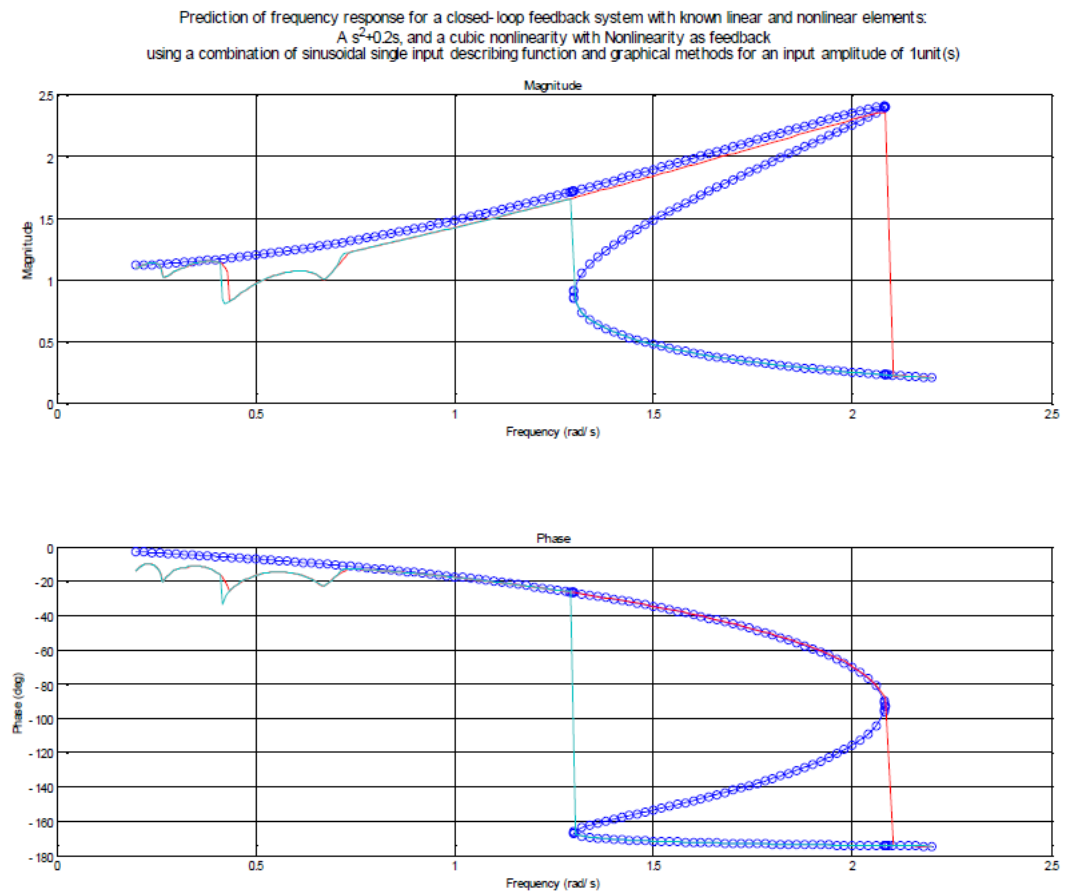
**Figure 5.11** Solutions from the graphical method for the frequency response with  $c = 0.2$

A software tool [5.14] was used to obtain sweep frequency response tests in Simulink. The tool was modified so that it could perform bi-directional sweeps necessary for the jump phenomenon investigation. The input sweep generator changes the frequency of the sinusoidal input from  $\omega_1$  to  $\omega_n$  rad/s progressively so that  $\omega_k = r^{-1} \omega_{k-1}$ ,  $r^{-1} > 1$  for sweep up, and from  $\omega_n$  to  $\omega_1$  so that  $\omega_{k-1} = r \omega_k$ ,  $r < 1$  for sweep down. The frequency increment is only applied after the output satisfies a steady state test. This is done by checking over the last  $n$  cycles of the output and confirming that the phase change is less than a specified tolerance,  $\epsilon$ . Averaging is also performed for  $m$  specified cycles of the output to compute the result. To obtain the results  $n$  and  $m$  were set to 3,  $\epsilon$  to  $10^{-5}$  and  $r$  to 0.995. When the increment in frequency is performed the system state conditions are preserved between frequencies, thus ensuring that conditions are similar to the previously mentioned analogue computer approach. The measured output is the fundamental amplitude of the output waveform determined by Fourier analysis, which should give best agreement with the DF analysis.

The results obtained for the cases of  $c = 0.4$  and  $0.2$  with  $R = 1$  are shown in Figures 5.12 and 5.13, superimposed on the theoretical curves. It can be seen that there is good agreement between the theoretical and experimental results particularly around the jump frequencies.



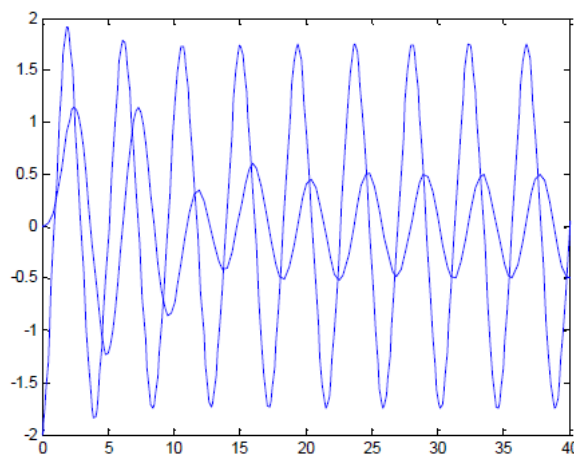
**Figure 5.12** Swept frequency response test for  $c = 0.4$  compared with the theory.



**Figure 5.13** Swept frequency response test for  $c = 0.2$  compared with the theory.

In addition spot tests at various frequencies within the predicted jump range were undertaken. This was done in Simulink by defining the input as a unit amplitude sine wave with phase zero and simulating  $G(s)$  as in Figure 5.9, which allows initial conditions to be applied to the integrator. For  $c = 0.4$  two almost sinusoidal output amplitude solutions in good agreement with the theory were obtained in the frequency range 1.38 to 1.49 as shown in Figure 5.14 for a frequency of 1.45. The smaller amplitude solution was obtained with a zero initial condition and the larger one with an initial condition of -2.





**Figure 5.14** Plots of the two outputs for  $c = 0.4$  and input frequency of 1.45 rad/s.

For  $c = 0.2$  the larger output amplitude solution could only be found with an initial condition on the integrator for frequencies above 1.54 up to 2.1 rad/s. With a zero initial condition on the integrator the output was nearly sinusoidal at the input frequency until a frequency just exceeding 1.66 rad/s was reached when the lower amplitude contained a significant third harmonic which remained until a frequency of around 1.75 rad/s. Between approximately 1.75 rad/s and 1.80 rad/s the third subharmonic disappeared and the output again became almost sinusoidal at the input frequency. The third subharmonic component again returned from just above 1.80 rad/s until it finally disappeared at 2.05 rad/s.

Please click the advert

# Brain power



By 2020, wind could provide one-tenth of our planet's electricity needs. Already today, SKF's innovative know-how is crucial to running a large proportion of the world's wind turbines.

Up to 25 % of the generating costs relate to maintenance. These can be reduced dramatically thanks to our systems for on-line condition monitoring and automatic lubrication. We help make it more economical to create cleaner, cheaper energy out of thin air.

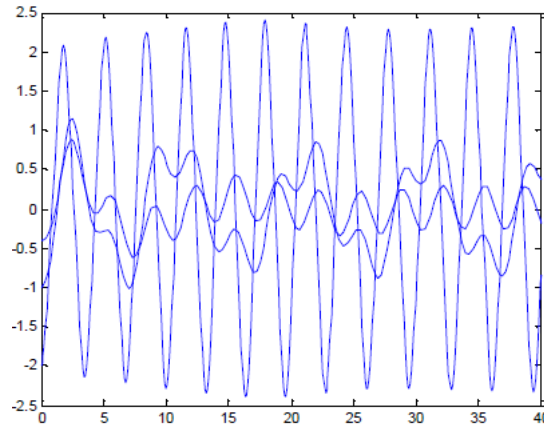
By sharing our experience, expertise, and creativity, industries can boost performance beyond expectations. Therefore we need the best employees who can meet this challenge!

The Power of Knowledge Engineering

Plug into The Power of Knowledge Engineering.  
Visit us at [www.skf.com/knowledge](http://www.skf.com/knowledge)



Further investigation for frequencies between 1.66 rads/s and 2.10 rads/s using different initial conditions on the integrator revealed three stable modes according to the initial conditions. These were two approximately sinusoidal amplitudes at the input frequency in good agreement with the theory and a third with a significant subharmonic content. Figure 5.15 shows these three outputs for an input frequency of 1.9 rads/s.



**Figure 5.15** Plots of the three outputs for  $c = 0.2$ ,  $R = 1$  and an input frequency of 1.9 rads/s.

This example clearly illustrates that use of the approximate DF method can determine steady state frequency response solutions but cannot predict what can occur in a particular situation, as this is determined by the transient dynamics when more than one stable state exists. The possibility of the third solution existing, namely the subharmonic oscillation can also be predicted by the DF method as is shown in the following.

If the input to the feedback loop is again taken as  $R \cos(\omega t + \beta)$ , then the input to the cubic nonlinearity, which is also the output, may be taken as  $a \cos \omega t + b \cos((\omega t/3) + \varphi)$  to investigate the possibility of a third subharmonic. The fundamental output signal from the cubic is  $aM_a(a|b)\cos(\omega t + \phi_a(a|b))$  and the loop equation for the fundamental, with the dependence of  $g$  and  $\theta$ , on  $\omega$  shown explicitly in the equation, is

$$Rg(\omega)\cos(\omega t + \beta + \theta(\omega)) - aM_a(a|b)g(\omega)\cos(\omega t + \theta(\omega) + \phi_a(a|b)) = a \cos \omega t \quad (5.55)$$

For the third subharmonic the corresponding loop equation is

$$-bM_b(a|b)g(\omega/3)\cos((\omega t/3) + \theta(\omega/3) + \phi_b(a|b) + \varphi) = b \cos((\omega t/3) + \varphi) \quad (5.56)$$

These two equations provide four nonlinear equations which can be solved for the unknowns  $a$ ,  $b$ ,  $\beta$  and  $\varphi$  to obtain a possible subharmonic solution when  $R$  and  $\omega$  are given. A quick test for a possible subharmonic can be obtained from the phase shift requirement of equation (5.56), namely

$$\theta(\omega/3) + \phi_b(a|b) = -180^\circ \quad (5.57)$$

From equation (5.2) it can be shown that

$$\phi_b(a|b) = -\tan^{-1} \left[ ab \sin 3\varphi / (b^2 + 2a^2 + ab \cos 3\varphi) \right] \quad (5.58)$$

Differentiation of the expression in the brackets shows that it is a maximum when  $\cos 3\varphi = -ab/(b^2 + 2a^2)$  which gives the maximum value for  $\phi_b(a|b)$  of  $\pm \sin^{-1} [ab/(b^2 + 2a^2)]$ . Denoting the ratio of  $b/a$  by  $\rho$  gives  $\sin^{-1} [\rho/(\rho^2 + 2)]$ , which on differentiation gives a maximum when  $\rho = \sqrt{2}$  and a maximum value of  $\sin^{-1}(\sqrt{2}/4) = 20.7^\circ$ .

Thus a subharmonic can only occur provided the phase lag of  $G(j\omega)$  is greater than  $\approx 159^\circ$ . The lowest frequency for which this occurs is approximately 0.52 rads/s, so that subharmonic oscillations might be possible when the input frequency exceeds 1.56 rads/s which is in reasonable agreement with the experimental results. The subharmonic only lasts for a small frequency range as the loop gain to the subharmonic of  $M_b(a|b)g(\omega/3)$  drops below unity. It can be shown that the inverse of the describing function for the subharmonic is a set of circles dependent on  $\rho$  within the sector of angle  $\pm 20.7^\circ$  about the negative real axis of the Nyquist plane. Thus for a subharmonic to exist for the cubic nonlinearity in a feedback loop the Nyquist plot of  $G(j\omega)$  must enter this sector.

The disagreement in Figures 5.12 and 5.13 at low frequencies is due to the poor filtering of the harmonics provided by  $G(j\omega)$ . If the cubic is moved to the forward loop in front of  $G(s)$  the discrepancy at low frequencies is less as the sinusoidal loop input signal passes straight to the cubic nonlinearity. This is illustrated by the frequency response shown in Figure 5.16 for the case of  $c = 0.4$  and  $R = 1$ . Here, however, another situation arises in that theoretically there are only two solutions above a frequency of around 2.25 rads/s. and thus there is no jump down on the upward frequency sweep as confirmed by the sweep frequency test, due to the loop having a negligible feedback signal. In practice, of course the increasing gain of the cubic is not possible and a jump down will result. Spot testing shows that if no initial conditions are placed on the integrator the solutions of the decreasing frequency sweep are obtained. However above the frequency of 2.25 rads/s the high amplitude stable solution can be obtained by putting initial conditions on the integrator.

Prediction of frequency response for a closed-loop feedback system with known linear and nonlinear elements:  
 $A s^2 + 0.4s$ , and a cubic nonlinearity with a Hammerstein forward path configuration  
 using a combination of sinusoidal single input describing function and graphical methods for an input amplitude of 1 unit(s)

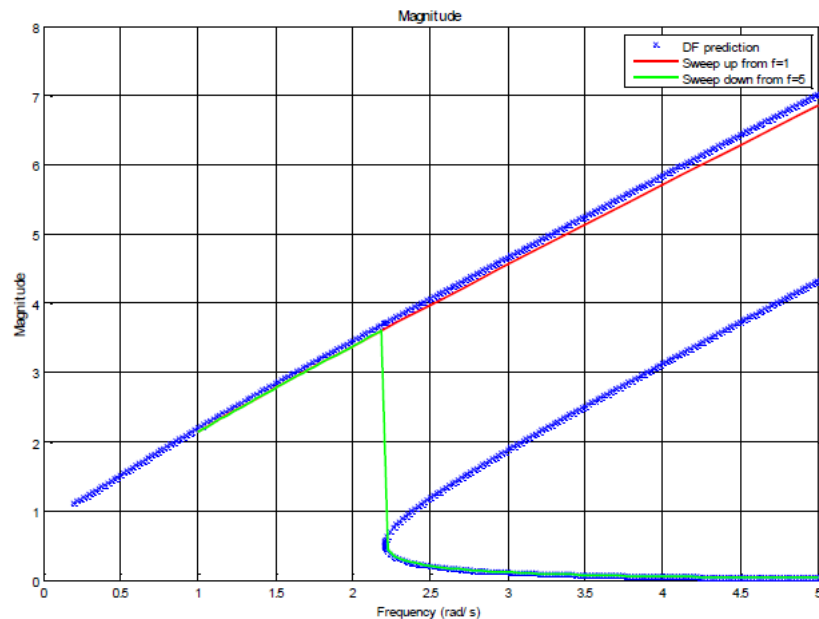


Figure 5.16 Frequency response for cubic in the forward path with  $c = 0.4$ .

Please click the advert

## Are you considering a European business degree?

**LEARN BUSINESS** at university level. We mix cases with cutting edge research working individually or in teams and everyone speaks English. Bring back valuable knowledge and experience to boost your career.

**MEET** a culture of new foods, music and traditions and a new way of studying business in a safe, clean environment – in the middle of Copenhagen, Denmark.

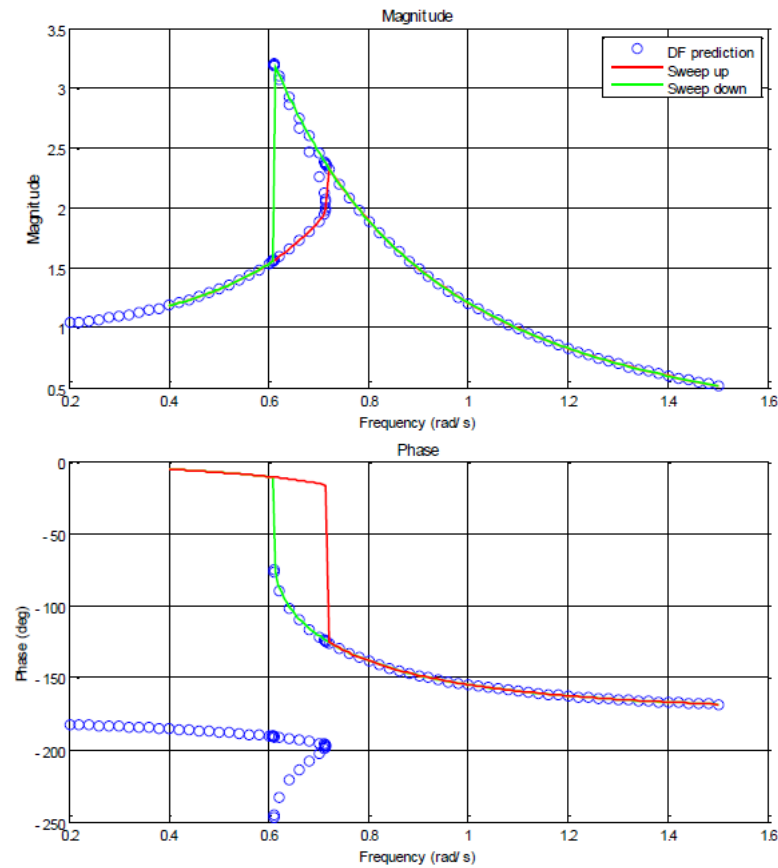
**ENGAGE** in extra-curricular activities such as case competitions, sports, etc. – make new friends among CBS' 18,000 students from more than 80 countries.



See what we look like  
and how we work on [cbs.dk](http://cbs.dk)



Results are given in Figure 5.17 for an ideal saturation characteristic in the forward path before the same  $G(s)$  with  $c = 0.2$  and  $R = 1$ . In this case it is seen that the theoretical resonant peak bends in a backward direction so that the smaller jump occurs on the upward frequency sweep.



**Figure 5.17** Frequency response results for saturation in the forward path,  $c = 0.2$  and  $R = 1$ .

## 5.8 Conclusions

In this chapter the response of a nonlinearity to two sinusoidal inputs was first discussed and although it was shown that a mathematical formulation is straightforward obtaining analytical solutions for other than simply defined mathematical characteristics is not. Thus numerical computation approaches have to be adopted for most practical situations. The concepts were used, however, to show how various properties of feedback systems could be investigated by considering two frequencies present in the loop. The analytical work has been supported by simulation results and the chapter was concluded by a detailed studied of jump resonance.



## 5.9 References

- 5.1 McNamara, OP & Atherton, DP 1985, An Iterative Method for Limit Cycle Determination, Control '85, Cambridge, pp 252-257.
- 5.2 Atherton, DP, McNamara, OP, Wadey, MD & Goucem A 1985, SUNS: the Sussex University Nonlinear Control Systems Software. 3rd IFAC/ICAD Symposium on Computer Aided Design in Control Engineering Systems, Copenhagen, pp 173-178.
- 5.3 McNamara, OP & Atherton, DP, 1987, Limit Cycle Prediction in Free Structured Nonlinear Systems, IFAC Congress, Munich, Volume 8, pp 23-28.
- 5.4 Atherton, DP & Ramani, N 1975, A Method for the Evaluation of Limit Cycles. International Journal of Control, Vol 21, pp 375-384
- 5.5 Hammond, PH 1958, Feedback Theory and its Application, Chapter 11, EUP, London.
- 5.6 West, JC and Douce, JL 1952, The Frequency Response of a Certain Class of Nonlinear Feedback Systems, British Journal of Applied Physics Vol 5, pp.204-210.
- 5.7 Singh, YP 1964, Graphical Method for finding the Closed Loop Frequency Response of Nonlinear Feedback Control Systems, Proc IEE, Vol 112, pp 2167-2170.
- 5.8 West, JC & Nikiforuk, P 1954, The Behaviour of a Remote Position Control Servomechanism with Hard Spring Nonlinear Characteristics, Proc IEE, Vol 101, part II, pp 481-487.
- 5.9 West, JC, Jayawant, BV & Rea, DP 1967, Transition Characteristics of the Jump Phenomenon in Nonlinear Resonant Circuits, Proc IEE, Vol 114, pp 381-392.
- 5.10 Lamba, SS & Kavanagh, RJ 1971, The phenomenon of Isolated Jump Resonance and its Applications, Proc IEE, Vol 118, pp 1047-1050.
- 5.11 Hirai, K & Sawai, N 1978, A General Criterion for Jump Resonance of Nonlinear Control Systems, IEEE Trans. on Auto. Control, pp. 896-901.
- 5.12 Hatanaka, H 1963, The Frequency Response and Jump Resonance Phenomena of Nonlinear Feedback Control Systems, Trans ASME J Basic Eng. D-85, pp236-242.
- 5.13 Wong, HK 2010, Private communications
- 5.14 Royalty, J 2007, Dynamic Signal Analyzer , MATLAB Central [online], <http://www.mathworks.com/matlabcentral/fileexchange/14611>


# 6 Limit cycles in relay systems

## 6.1 Introduction

In the early days of control engineering, when power electronics was primitive compared with the situation today, relays were commonly used as they provided a cheap form of power amplification. They are still used in temperature control, where the temperature is controlled to limit cycle around the desired value, and switching techniques using modern power electronics are also frequently used. The distinction between switching nonlinearities and other nonlinearities is that their input does not control the output continuously but only determines the instant of switching. The input has no control of the output between switching instants. This feature, that the input has no control of the output between switching instants, allows a unique theory to be developed for analysing limit cycles in feedback loops with relay elements.

The starting point for the analysis is to assume a specific output waveform from the relay and develop equations which must be satisfied for the assumed waveform to exist. The most common situation is the case of a relay with no dead zone where the output will typically be assumed to be a square waveform with a one to one mark space ratio. More complex modes with multiple pulses per cycle can exist in some systems and their prediction is still possible, however more nonlinear equations need to be satisfied, thus typically requiring a computational approach. The analysis starts by assuming a periodic output waveform from the relay and then the requirement is to obtain the steady state periodic output this produces after passage through the transfer function  $G(s)$ . This is also a problem of interest in circuit theory [6.1].

Please click the advert



**The financial industry needs a strong software platform**  
That's why we need you

SimCorp is a leading provider of software solutions for the financial industry. We work together to reach a common goal: to help our clients succeed by providing a strong, scalable IT platform that enables growth, while mitigating risk and reducing cost. At SimCorp, we value commitment and enable you to make the most of your ambitions and potential.

Are you among the best qualified in finance, economics, IT or mathematics?

**Find your next challenge at**  
[www.simcorp.com/careers](http://www.simcorp.com/careers)



[www.simcorp.com](http://www.simcorp.com)  
MITIGATE RISK | REDUCE COST | ENABLE GROWTH

For application to the relay feedback problem three different methods have been proposed for doing the analysis all of which produce identical results. A time domain approach was developed by Hamel [6.2], a frequency domain approach by Tsytkin [6.3], and a time domain approach using a state space representation by Chung [6.4, 6.5]. The frequency domain approach uses a 'summed' frequency locus which makes it easy to compare with the DF method so that this will be the method considered below. The formulation is done in terms of  $A$  loci which are more general than the Tsytkin loci, and this is followed with details on the properties of  $A$  loci and their evaluation. Solutions for limit cycles of the 'fundamental' form are discussed. In section 6.5 it is shown that an exact condition can be found giving necessary and sufficient conditions for the stability of any predicted limit cycle, which is an extremely useful result. This enables the study in section 6.6 of some very interesting problems relating to limit cycles in relay systems which have been obtained from software implementations of the theory presented. Topics discussed include limit cycles with multiple pulses per period, limit cycles with sliding and concepts for the prediction of chaotic motion. Finally in section 6.7 the harmonically forced response of relay systems is considered.

## 6.2 The Frequency Domain Approach

Consider a periodic pulse waveform for which a one period example is illustrated in Figure 6.1, where it is assumed to be odd symmetrical and to have eight pulses in a period. If any one of the eight pulses is denoted by  $y_i(t)$  which is assumed to be of width  $\Delta t_i$  and to start at time  $t_i$  then it is straightforward to show that it has the Fourier series

$$y_i(t) = (h\Delta t_i/T) + (h/\pi) \sum_{n=1}^{\infty} (1/n) [\sin n\omega\Delta t_i \cos n\omega(t - t_i) + (1 - \cos n\omega\Delta t_i) \sin n\omega(t - t_i)] \quad (6.1)$$

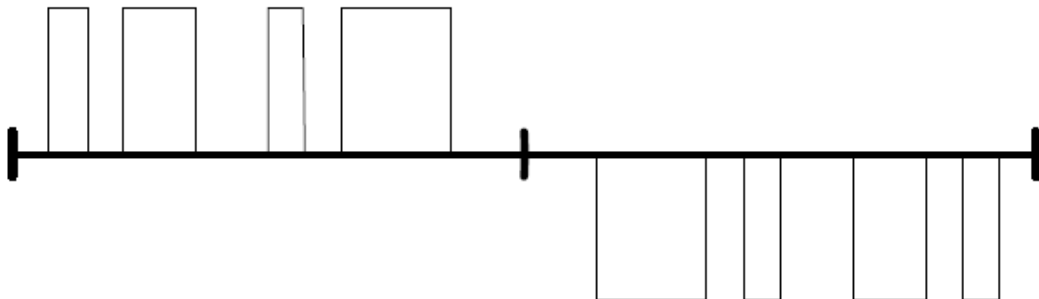
If  $y_i(t)$  is the input to a linear system with the transfer function  $G(s)$ , then the Fourier series for the output  $c_i(t)$ , assuming  $\lim_{s \rightarrow \infty} G(s) = 0$  and  $G(0)$  is finite, is given by

$$c_i(t) = (h\Delta t_i G(0)/T) + (h/\pi)$$

$$\sum_{n=1}^{\infty} (g_n/n) \{ \sin n\omega\Delta t_i \cos [n\omega(t - t_i) + \varphi_n] + (1 - \cos n\omega\Delta t_i) \sin [n\omega(t - t_i) + \varphi_n] \} \quad (6.2)$$

where the frequency response of  $G(s)$  is denoted by

$$G(jn\omega) = g_n e^{j\varphi_n} = U_G(n\omega) + jV_G(n\omega) \quad (6.3)$$



**Figure 6.1** One period of an odd symmetrical multipulse waveform.



If equation (6.2) is differentiated then the Fourier series for  $\dot{c}_i(t)$ , provided  $\lim_{s \rightarrow \infty} sG(s) = 0$ , is

$$\dot{c}_i(t) = (\omega h/\pi)$$

$$\sum_{n=1}^{\infty} (g_n/n) \left\{ -\sin n\omega\Delta t_i \sin[n\omega(t - t_i) + \varphi_n] + (1 - \cos n\omega\Delta t_i) \cos[n\omega(t - t_i) + \varphi_n] \right\} \quad (6.4)$$

When  $\lim_{s \rightarrow \infty} sG(s) \neq 0$ ,  $c_i(t)$ , will have discontinuities at the switching instants  $t_i$  and  $t_i + \Delta t_i$ . Since the Fourier series tends to the mean value of the function at a discontinuity at time,  $t$ , then a term  $0.5\{y_i(t^+) - y_i(t^-)\} \lim_{s \rightarrow \infty} sG(s)$  with appropriate sign must be included in equation (6.4) to obtain the correct value of  $c_i(t)$  at these two time instants. It is convenient to express  $c(t)$  and  $\dot{c}(t)$  in terms of the A locus [6.6, 6.7],  $A(\theta, \omega)$ , where  $\theta = \omega t$ , and is defined for a transfer function  $G(s)$  by

$$A_G(\theta, \omega) = \text{Re } A_G(\theta, \omega) + j \text{Im } A_G(\theta, \omega) \quad (6.5)$$

with

$$\text{Re } A_G(\theta, \omega) = \sum_{n=1}^{\infty} V_G(\theta, n\omega) \sin n\theta + U_G(\theta, n\omega) \cos n\theta \quad (6.6)$$

and

$$\text{Im } A_G(\theta, \omega) = \sum_{n=1}^{\infty} (1/n) [V_G(\theta, n\omega) \cos n\theta + U_G(\theta, n\omega) \sin n\theta] \quad (6.7)$$

$A_G(\theta, \omega)$  is a generalised summed frequency locus with its real and imaginary values at a particular frequency,  $\omega$ , depending on weighted values, according to the choice of,  $\theta$ , of the real and imaginary values of  $G(j\omega)$  at the frequencies  $n\omega$  for  $n = 1, 2, \dots, \infty$ . In particular for  $\theta = 0$  the real (imaginary) part of  $A_G(\theta, \omega)$  depends only upon the real (imaginary) part of  $G(j\omega)$  at frequencies  $n\omega$ . It is possible to evaluate  $\text{Re } A_G(\theta, \omega)$  and  $\text{Im } A_G(\theta, \omega)$  in closed form for specific transfer functions from known infinite series, as is discussed in the next section, or approximately by computationally summing to a finite number,  $n$ , of terms.

Using the above relationships in equations (6.2) and (6.4) gives

$$c_i(t) = (h\Delta t_i G(0)/T) + (h/\pi) [\text{Im } A_G(-\omega t + \omega t_i, \omega) - \text{Im } A_G(-\omega t + \omega t_i + \omega\Delta t_i, \omega)] \quad (6.8)$$

and

$$\dot{c}_i(t) = (\omega h/\pi) [\text{Re } A_G(-\omega t + \omega t_i, \omega) - \text{Re } A_G(-\omega t + \omega t_i + \omega\Delta t_i, \omega)] \quad (6.9)$$

Equation (6.9) may also be seen to be obtainable directly from equation (6.8) since from the definitions of equations (6.6) and (6.7) it can be seen that

$$\frac{d \text{Im} [A_G(\theta, \omega)]}{d\theta} = -\text{Re } A_G(\theta, \omega) \quad (6.10)$$

Any periodic pulse waveform can be constructed from a summation of waveforms  $y_i(t)$  with different origins and pulse widths. For example, the symmetrical square wave,  $y(t)$ , of period  $T$  is given by

$$y(t) = \sum_{i=1}^2 y_i(t) \quad (6.11)$$

with  $y_1(t)$  switching positive at time  $t_1$ ,  $y_2(t)$ , switching negative at time  $t_2$ ; and with  $t_2 = t_1 + 0.5T$  and  $\Delta t_1 = \Delta t_2 = 0.5T$ .

Choosing the time origin  $t_1 = 0$ , then the output  $c(t)$  becomes

$$c(t) = (2h/\pi) [\text{Im} A_G(-\omega t, \omega) - \text{Im} A_G(-\omega t + 0.5\omega T, \omega)] \quad (6.12)$$

Since  $0.5\omega T = \pi$ , it is easily shown that this gives

$$c(t) = (4h/\pi) [\text{Im} A_G^o(-\omega t, \omega)] \quad (6.13)$$

and, similarly,

$$\dot{c}(t) = (4\omega h/\pi) [\text{Re} A_G^o(-\omega t, \omega)] \quad (6.14)$$

Here the  $A$  loci  $A_G^o$  are again defined by equations (6.5) to (6.7) but with the summations including odd terms only, that is  $n = 1, 3, 5, \dots, \infty$ . Before proceeding to show how limit cycles can be found using these expressions it is appropriate to summarise some properties of the  $A$  loci and show how they can be obtained for specific transfer functions.

Please click the advert



What do you want to do?

No matter what you want out of your future career, an employer with a broad range of operations in a load of countries will always be the ticket. Working within the Volvo Group means more than 100,000 friends and colleagues in more than 185 countries all over the world. We offer graduates great career opportunities – check out the Career section at our web site [www.volvogroup.com](http://www.volvogroup.com). We look forward to getting to know you!

**VOLVO**  
AB Volvo (publ)  
[www.volvogroup.com](http://www.volvogroup.com)

VOLVO TRUCKS | RENAULT TRUCKS | MACK TRUCKS | VOLVO BUSES | VOLVO CONSTRUCTION EQUIPMENT | VOLVO PENTA | VOLVO AERO | VOLVO IT  
VOLVO FINANCIAL SERVICES | VOLVO 3P | VOLVO POWERTRAIN | VOLVO PARTS | VOLVO TECHNOLOGY | VOLVO LOGISTICS | BUSINESS AREA ASIA

Download free ebooks at [bookboon.com](http://bookboon.com)

### 6.3 Properties and Evaluation of A loci

Some further properties of the A loci, in addition to equation (6.10), which are easily found from the definitions are:-

- (1) The  $A^\circ$  locus for  $\theta = 0$  is identical with the Tsytkin locus [6.3],  $\Lambda(\omega)$ , if  $\lim_{s \rightarrow \infty} sG(s) = 0$ , apart from a constant factor. The relationship is

$$\Lambda(\omega) = (4h/\pi)A^\circ(0, \omega) \quad (6.15)$$

The  $A^\circ$  locus can thus be regarded as a generalised Tsytkin locus.

- (2) The A locus satisfies the superposition property, that is if the linear plant  $G(s) = G_1(s) + G_2(s)$ , then

$$A_G(\theta, \omega) = A_{G_1}(\theta, \omega) + A_{G_2}(\theta, \omega) \quad (6.16)$$

- (3) If  $G_1(s) = G(s)e^{-s\tau}$ , then

$$A_{G_1}(\theta, \omega) = A_G(\theta + \omega\tau, \omega) \quad (6.17)$$

- (4) The loci are periodic in  $\theta$ , with period  $2\pi$ , that is

$$A_G(\theta, \omega) = A_G(\theta + 2\pi, \omega) \quad (6.18)$$

- (5) For  $A^\circ$  the periodicity is odd, that is

$$A_G^\circ(\theta, \omega) = -A_G^\circ(-\theta, \omega) \quad (6.19)$$

Since any plant transfer function can be written in terms of a summation of transfer functions having a real or complex pair of poles, use of equation (6.16) allows A loci for most transfer functions to be obtained in terms of the A loci of the few basic transfer functions given in Table 6.1 in the Appendix 6.10. In addition A loci for transfer functions with time delays can be obtained using equation (6.17). Expressions for the A loci of the transfer functions of Table 6.1 are given in Tables 6.2 and 6.3 respectively, for the  $A^\circ$  and A loci. For the real pole situations the loci are expressible in terms of well known sine and cosine series.

## 6.4 Solving for Limit Cycles

The results of the previous section have shown how a closed form expression can be obtained for the output of a linear transfer function with a periodic pulse waveform input. Thus the signal fed back to the relay input can be found. If this pulse type limit cycle is to exist then the following three conditions need to be satisfied:-

- (1) The input must pass through the relay switching values in the correct direction at the appropriate times to provide the assumed periodic output. The number of these 'switching conditions' to be satisfied will depend on the form of periodic mode being investigated. The simplest case is discussed in the next section for a relay with no dead zone with a symmetrical square wave output, which yields one switching condition and thus the requirement to solve one nonlinear algebraic equation. For complex waveforms there may be many switching conditions resulting in the requirement to solve many nonlinear algebraic equations
- (2) Once the switching conditions have been found the relay input waveform can be computed. This must satisfy the 'continuity conditions', that is it must not pass through the relay switching levels at times which would cause the relay output to be different from that assumed.
- (3) The limit cycle must be stable.

Although analytical solutions can be obtained for some simple cases of one or two switchings per limit cycle period, the best approach is to use software with graphics [6.8]. This then allows the above three conditions to be accomplished, namely solution of the nonlinear algebraic equations of (1), display of solution waveforms to check (2), and the stability computations, discussed in section 6.5, to be carried out for (3)

### 6.4.1 Relay with no Dead Zone

For a relay with no dead zone the simplest assumption for the periodic output is a square wave with 1:1 mark space ratio. Equations (6.13) and (6.14) give expressions for the waveform and its derivative produced at the output of the linear transfer function  $G$ , where the positive switching of the square wave was assumed to be at  $t = 0$ . Taking the input to the relay,  $x(t) = -c(t)$ , then the required switching condition, assuming the relay to have a hysteresis of  $\pm\Delta$ , that is it switches from  $-h$  to  $+h$  for an input of  $\Delta$  and from  $+h$  to  $-h$  for an input of  $-\Delta$ , is

$$x(0^-) = \Delta \text{ and } \dot{x}(0^-) > 0 \quad (6.20)$$

The  $-$  is used to show that the time should be taken before any possible discontinuity which may be created due to the switching. Thus using (6.13) and (6.14) one obtains

$$\text{Im}A_o^o(0, \omega) = -\pi\Delta/4h \text{ and } \text{Re}A_o^o(0, \omega) < 0 \quad (6.21)$$

provided  $\lim_{s \rightarrow \infty} sG(s) = 0$ .

For a given  $G(s)$  the solution for the unknown frequency of the limit cycle,  $\omega$ , can be found by solving the equality of equation (6.21) and checking the inequality, or done graphically from a plot of  $A_G^o(0, \omega)$  on a Nyquist diagram to obtain where it intersects the line  $-\pi\Delta/4h$ . This approach allows easy comparison with the DF method where the solution is the intersection of  $G(j\omega)$  with the same line. Note from the definition  $A_G^o(0, \omega)$  has real and imaginary parts given by

$$\operatorname{Re} A_G^o(0, \omega) = \sum_{n=1(2)}^{\infty} U_G(0, n\omega) = \sum_{n=1(2)}^{\infty} \operatorname{Re} G(jn\omega) \quad (6.22)$$

and

$$\operatorname{Im} A_G^o(0, \omega) = \sum_{n=1(2)}^{\infty} (1/n) V_G(0, n\omega) = \sum_{n=1(2)}^{\infty} (1/n) \operatorname{Im} G(jn\omega). \quad (6.23)$$

Thus the locus can be obtained approximately from these expressions by taking a small value for  $n$ , and exactly by summing to a large value of  $n$  computationally or using the analytical results of Table 6.2.

For the specific example of  $G(s) = K/[s(1 + s\tau)]$ , one has using Table 6.2 that

$$\operatorname{Im} A_G^o(0, \omega) = (K\pi\tau/4)[(-\pi/2\lambda) + \tanh(\pi/2\lambda)] \quad (6.24)$$

where  $\lambda = \omega\tau$ . Thus using equation (6.21) the solution for the limit cycle is

$$(\pi/2\lambda) - \tanh(\pi/2\lambda) = \Delta h K / \tau \quad (6.25)$$

and the inequality is satisfied as  $\operatorname{Re} A_G^o(0, \omega) < 0$  for all  $\omega$ .

The DF method for this problem yields the equation

$$\lambda(1 + \lambda^2) = 4hK\tau/\Delta \quad (6.26)$$

for the normalised frequency,  $\lambda$ , of the limit cycle.

For the particular case of  $K = 1$ ,  $\tau = 1$ ,  $h/\Delta = 3$ , the exact limit cycle frequency solution is 1.365 and that given by the DF method is 1.352. As expected if the solution is examined as the parameter,  $\tau$ , changes then, since as  $\tau$  reduces the transfer function is a worse low pass filter, the DF solution becomes less accurate but it is less than 3% in error for  $\tau > 0.25$ .

When the square wave is asymmetrical, that is not a 1:1 mark to space ratio, the two pulse trains of equation (6.11) will be  $y_1(t)$  switching positive at  $t_1 = 0$  with  $\Delta t_1 = \Delta t$ , and  $y_2(t)$  switching negative at time  $t_2 = \Delta t$  and with  $\Delta t_2 = T - \Delta t$ . The required relay inputs of  $\Delta$  and  $-\Delta$  at the switching times of  $0^-$  (positive) and  $\Delta t^-$  (negative) then yield two nonlinear algebraic equations to solve for  $\Delta t$  and  $\omega$ .

### 6.4.2 Relay with Dead Zone

To evaluate any odd symmetrical limit cycle in a loop having a relay with dead zone, the relay output can again be assumed to be the sum of two pulse trains as given in equation (6.11) but now with pulse widths  $\Delta t_1 = \Delta t_2 = \Delta t \neq 0.5T$ . The positive switching condition at time  $t_1$ , which will be taken as 0, will be

$$x(0^-) = \delta + \Delta \text{ and } \dot{x}(0^-) > 0 \quad (6.27)$$

and the negative one at time  $\Delta t$

$$x(\Delta t^-) = \delta - \Delta \text{ and } \dot{x}(\Delta t^-) > 0 \quad (6.28)$$

The expressions for  $c(t)$  and  $\dot{c}(t)$  can be shown to be

$$c(t) = (2h/\pi)[\text{Im}A_o^o(0, \omega) - \text{Im}A_o^o(\omega\Delta t, \omega)] \quad (6.29)$$

and

$$\dot{c}(t) = (2\omega h/\pi)[\text{Re}A_o^o(0, \omega) - \text{Re}A_o^o(\omega\Delta t, \omega)]. \quad (6.30)$$

Please click the advert

*RedStarResume.*

## Do you want your Dream Job?

More customers get their dream job by using RedStarResume than any other resume service.

RedStarResume can help you with your job application and CV.

**Go to: [Redstarresume.com](http://Redstarresume.com)**

**Use code "BOOKBOON" and save up to \$15**

(enter the discount code in the "Discount Code Box")

Thus that the switching conditions, for  $x(t) = -c(t)$ , are

$$\begin{aligned} \operatorname{Im} A_G^o(0, \omega) - \operatorname{Im} A_G^o(\omega \Delta t, \omega) &= -(\pi/2h)(\delta + \Delta) \\ \operatorname{Re} A_G^o(0, \omega) - \operatorname{Re} A_G^o(\omega \Delta t, \omega) &< 0 \end{aligned} \quad (6.31)$$

and

$$\begin{aligned} \operatorname{Im} A_G^o(0, \omega) - \operatorname{Im} A_G^o(\omega \Delta t, \omega) &= -(\pi/2h)(\delta - \Delta) \\ \operatorname{Re} A_G^o(0, \omega) - \operatorname{Re} A_G^o(\omega \Delta t, \omega) &> 0 \end{aligned} \quad (6.32)$$

For a specific example one thus has two nonlinear algebraic equations (6.31) and (6.32) to solve for the two unknowns  $\omega$  and  $\Delta t$ . Once these are known the pulse waveform is known and the limit cycle waveform can be obtained from the A locus expression or by other means. Plotting this waveform then allows condition (3) in section (6.4) to be checked. To investigate the possibility of an asymmetrical limit cycle for the relay with dead zone requires the solution of four nonlinear algebraic equations.

## 6.5 Limit Cycle Stability

When discussing limit cycle stability using the DF method the perturbation equation (4.5), which has periodically time varying coefficients, was found and is repeated below.

$$q(D)\delta x(t) + p(D)n'(x^*(t))\delta x(t) = 0 \quad (6.33)$$

In contrast to the DF case the exact limit cycle solution not a sinusoidal approximation is given by the above analysis for feedback loops with relays. Further if this equation is written in the state space form

$$\delta \dot{x}^*(t) = A(t)\delta x^*(t) \quad (6.34)$$

then it can be shown that the periodically time varying matrix,  $A(t)$ , is given by

$$A(t) = A + Bn'(x^*(t))C \quad (6.35)$$

where  $(A, B, C)$  is a state space description of  $G(s)$ . The matrix of equation (6.35) is piecewise constant and equal to  $A$  except over the infinitesimal time intervals when the relay switches. The value over the relay switching time may be calculated by taking the relay as a limiting case of saturation.

For example if the relay switches from  $-h$  to  $+h$  this is the limit of an ideal saturation characteristic as  $\varepsilon \rightarrow 0$  moving from a level  $-h$  at  $-\varepsilon$  to  $+h$  at  $+\varepsilon$  and the value of the second term in equation (6.35) is

$$\lim_{\varepsilon \rightarrow 0} BC(h/\varepsilon) = \lim_{\varepsilon \rightarrow 0} BC[2h/\dot{x}(t_i)\Delta] \quad (6.36)$$

where  $t_i$  is the time at which switching takes place and  $\Delta$  the duration of the switching time which tends to zero as  $\epsilon$  tends to zero. Thus the product  $A(t)\Delta$  over the infinitesimal switching time is given by

$$A(t)\Delta = BC[2h/\dot{x}(t_i)] \quad (6.37)$$

The term in the square brackets is the change in relay output divided by the derivative of the relay input at the switching instant. The value of this result is that a necessary and sufficient result, for the stability of a differential equation with a periodically time varying  $A$  matrix which is piecewise constant, is known [6.9, 6.10]. The result is that the system is stable if all the eigenvalues of the matrix  $Q$  have magnitude less than unity apart from one, corresponding to the periodic solution, which will have unit magnitude. The matrix,  $Q$ , is given by

$$Q = \prod_{i=1}^m A_i \Delta t_i \quad (6.38)$$

where the  $A_i$  are the constant values of the  $A$  matrix over the time periods  $\Delta t_i$  and  $T = \sum_{i=1}^m \Delta t_i$

is the limit cycle period with  $m$  values of the  $A$  matrix. For a symmetrical odd limit cycle  $Q$  may be computed over  $0.5T$ . Thus for a limit cycle in a feedback system with an ideal relay with levels  $\pm h$  switching at times 0 and  $0.5T$ , the  $Q$  matrix is given by

$$Q = \exp[AT/2] \exp[2hBC/\dot{x}(0)] \quad (6.39)$$

## 6.6 Some Interesting Limit Cycle Problems

For most feedback systems with relay elements the limit cycles obtained are of what might be called the ‘fundamental’ type discussed in section 6.4. That is (1) for a relay with no dead zone the relay will switch twice per period of the limit cycle, namely at times 0 and  $\Delta t$  in the period,  $T$ , where if the limit cycle has odd symmetry  $\Delta t = 0.5T$  and (2) for a relay with dead zone the relay will switch four times per period. In this section, rather than consider examples for these fundamental cases some more interesting examples are considered. The results for these problems were found using the theory described above and some extensions. The solutions were obtained using Fortran programs with interactive graphics to allow the display of the limit cycle waveforms.

### 6.6.1 Example 1 – Invalid Continuity Condition

Consider an ideal relay with outputs  $\pm 1$  in a negative feedback loop with the transfer function  $G(s) = (1 + 2s)/(s^2 + 1)(s + 1)$ , [6.10-6.13]. The Nyquist locus of this transfer function starts in the first quadrant and moves out to infinity at unit frequency due to the undamped second order term in the denominator. After a phase change of  $-180^\circ$  around an infinite semicircle it returns to the origin in the third quadrant. There is no intersection of the negative real axis so the describing function indicates no limit cycle. If the Tsypkin method is used then it can be shown that it gives limit cycle solution frequencies of  $\omega = 1/(2n + 0.5)$  for  $n = 1, 2, \dots, \infty$ . Checking the solution waveforms for these frequencies they are found to cross the zero switching level of the relay at times other than those assumed, due to the resonant nature of the plant transfer function. Thus they fail the continuity condition (2) given in section 6.4 and as predicted by the DF no limit cycle exists.



## 6.6.2 Example 2 – Multipulse Limit Cycles

Here a very interesting practical example concerning the possibility of multipulse oscillations in the attitude control system of a communication satellite in a geostationary orbit is discussed [6.14-6.17]. The roll-yaw loop is controlled by a pseudo-rate controller which consists of jet thrusters with feedback through a time constant which has a different value when the thrusters are on to when they are off. Thus on its own the controller is a relay feedback loop containing a mode dependent transfer function. The satellite has solar panels for power collection which means that the plant transfer function consists of the rigid body dynamics of a double integrator plus lightly damped second order transfer functions, corresponding to the panel's resonant modes, in parallel. A block diagram for the system is shown in Figure 6.2.

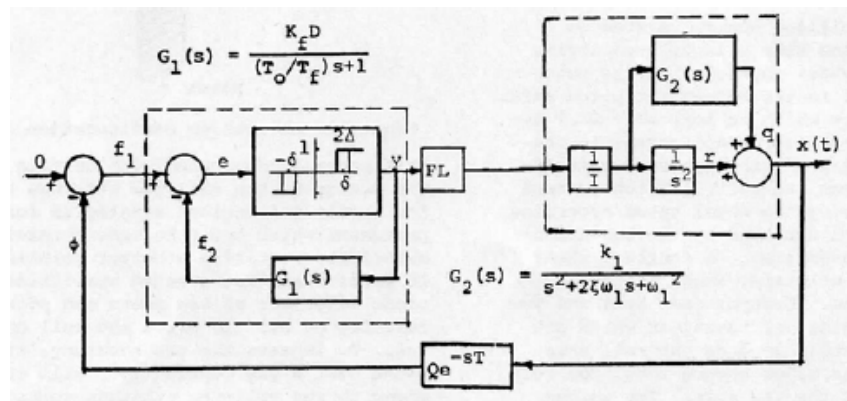


Figure 6.2 Block diagram for system of example 2 (copied from reference 6.15)

Try this...



The sequence 2, 4, 6, 8, 10, 12, 14, 16, ... is the sequence of even whole numbers. The 100th place in this sequence is the number...?

Challenging? Not challenging? Try more >>

[www.alloptions.nl/life](http://www.alloptions.nl/life)

Download free ebooks at [bookboon.com](http://bookboon.com)

Reference 6.15 discusses the determination of the multipulse limit cycles using the DF, but here the comments will refer to the use of the Tsytkin method discussed in papers [6.15, 6.16]. There are two original aspects in these papers the first being the extension of the Tsytkin method to evaluate quite complex limit cycles with multiple pulses per period and the second being an extension of the method to find limit cycles in mode dependent relay systems. Here the term mode dependent is used to refer to the fact that the loop transfer function changes with the relay output level. This topic is considered further in reference 6.17.

To determine the possible existence of an odd symmetrical multipulse oscillation of  $n$  pulses per half period requires a solution to be found to  $2n$  algebraic equations. Results were obtained by this approach in reference 6.16 for one assumed flexible mode and are given for the pseudo rate controller with equal and, then the usual case, unequal feedback time constants. Results with up to five pulses per half period were calculated and checked with analogue simulation results.

### 6.6.3 Example 3 – Limit Cycle with a Sliding Mode

The investigation of the step response of a relay control system in section 2.3.2 showed that a sliding motion, where the relay switches rapidly, was found. This therefore raises the interesting question of whether a limit cycle can exist with a sliding mode and if so how can it be determined theoretically. This problem is discussed in references 6.10, 6.18 and 6.19 where a relay with a dead zone of  $\pm 1$ , output levels  $\pm 1$  and linear transfer function  $G(s) = 8(s - 0.5)/(s^2 - 0.5s + 1)$  was studied. The relative degree of the transfer function was taken as one since it can be shown that sliding motion under relay control is not possible for a relative degree greater than unity. The Tsytkin method gave the solution for a limit cycle shown in Figure 6.3, which violates the continuity condition. The interesting point is that the waveform reverses immediately after the switching at A and then soon reverses again to pass through the switching level at C, which is argued is an indication of sliding with the first estimate for the time of sliding being from A to C.

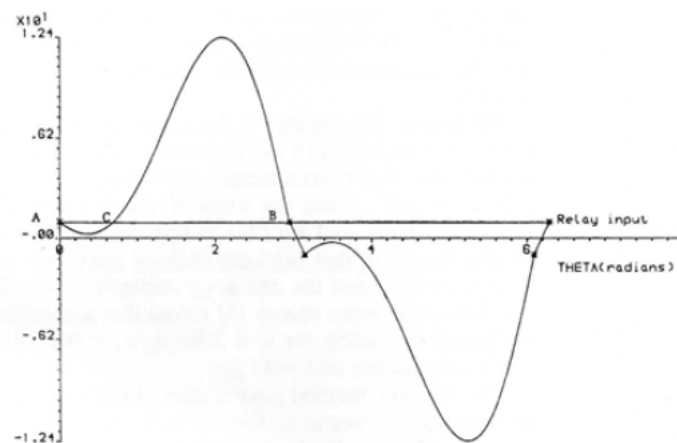


Figure 6.3 Tsytkin solution for limit cycle (taken from reference 6.19)

To see if a sliding solution could be obtained the relay output was approximated by a straight line over this period and the Tsytkin method extended to obtain a solution for a relay output waveform having this shape. A new limit cycle solution was then found showing that after the switching the calculated relay input remained almost constant as required to verify sliding. By adding a further linear segment to approximate the average relay output waveform in the sliding mode by two linear segments the method gave the limit cycle shown in Figure 6.4 with a frequency of 0.386 rads/s, which was in agreement with the simulation result. It can be seen that the relay input is now almost constant at the switching level during the sliding period.

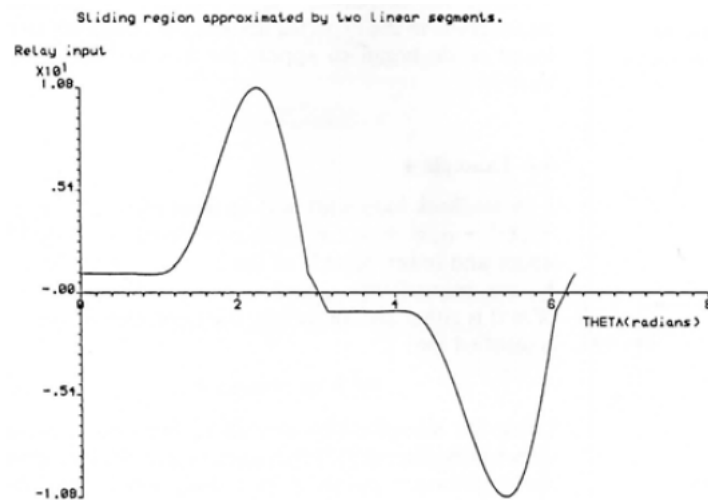


Figure 6.4 Limit cycle solution using Tsytkin extension (taken from reference 6.19)

The example has also been done in Simulink using a fixed step integration algorithm since the default option with a variable step length stops when sliding starts. The Simulink diagram is shown in Figure 6.5, where the blocks in the upper left hand corner generate a time vector,  $z$ , of equal length to the output vectors  $x$  and  $y$ .

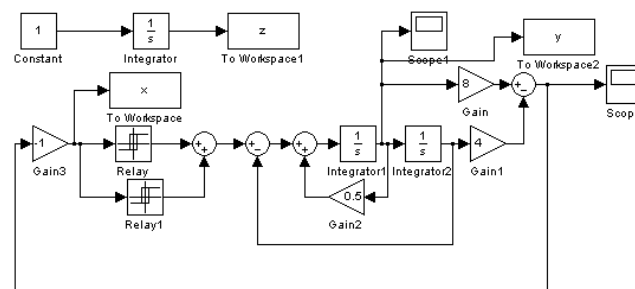


Figure 6.5 Simulink block diagram for system having limit cycle with sliding

The limit cycle at the input to the relay,  $x$ , is shown in Figure 6.6 and in the  $x$ - $y$  plane in Figure 6.7. The initial condition was set slightly off the limit cycle.

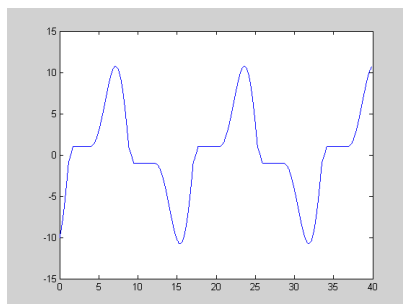


Figure 6.6. Waveform of limit cycle at the input to the relay.

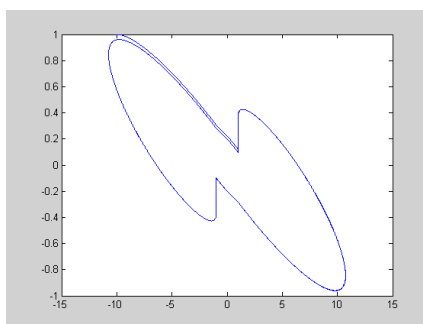



Figure 6.7 Limit cycle plot in the phase plane

Please click the advert

The iText logo, featuring a stylized blue 'i' and 'T' with an orange dot above the 'i', set against a background of overlapping light blue papers.

THIS **ebook** IS PRODUCED WITH **iText®**

### 6.6.4 Example 4 – Multiple Limit Cycle Solutions

For a feedback loop containing a single valued nonlinearity, which has a  $C(a)$  locus on the negative real axes and a linear transfer function which has a frequency response with more than one intersection of this axis, various possible limit cycles may result. A study of this situation using the approximate DF analysis, where both cubic and ideal relay nonlinearities are taken, can be found in reference 6.20. The Tsytkin method enables the existence of ‘fundamental’ limit cycles and their stability to be found but not limit cycles at combinations of the intersection frequencies. To illustrate the situation the case of an ideal relay in a negative feedback loop with the linear transfer function  $G(s) = 12(s+1)^2/s^3(s^2 + 0.3\lambda s + \lambda^2)$  as considered in reference 6.20 is discussed. The Nyquist locus for the transfer function with  $\lambda = 4.0$  is shown in Figure 6.8. The negative real axis is cut at the two frequencies  $\omega_1$  and  $\omega_2$ , with  $\omega_1 < \omega_2$ , and the ratio  $\alpha = |G(j\omega_1)|/|G(j\omega_2)|$  decreases as  $\lambda$  decreases. Using the IDF approach of the DF method the limit cycle at frequency  $\omega_1$  is always unstable and that at  $\omega_2$  is stable if  $\alpha > 2$  which corresponds to  $\lambda > 4.05$ . This compares with the Tsytkin solution which gave an unstable limit cycle at frequency  $\omega_1$  for  $\lambda > 3.73$ , but no solution for smaller values of  $\lambda$ . The limit cycle at  $\omega_2$  was only stable for  $\lambda > 3.73$ . Simulations gave good agreement with the Tsytkin theory for a single mode limit cycle but it cannot predict the combined mode. Using the SSDF and the ISSDF, the incremental describing function in the presence of two sinusoids, a stable combined frequency limit cycle was found to exist for  $\lambda < 4.05$  in reference 6.20. Good agreement was also found with the amplitude and frequency components in the combined mode.

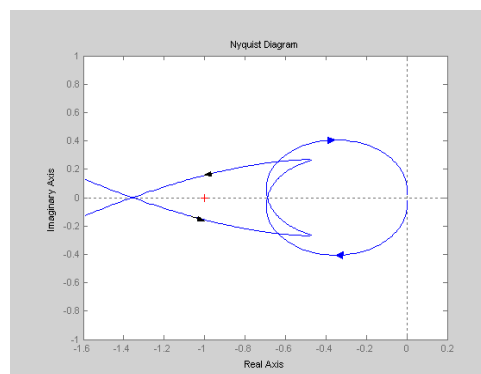


Figure 6.8 Nyquist plot for example 4

### 6.6.5 Example 5 – Chaotic Motion

Although no theoretical solution can be found for a chaotic motion there has been significant research done in trying to find conditions for its existence in autonomous feedback loops. In these studies both the approximate describing function approach [6.21] and Tsytkin method [6.22] have been used. Useful starting points are provided by the conjectures of Ogorzalek [6.23] and from the results of reference 6.21. Ogorzalek conjectured that for a negative feedback system with

$$G(s) = -\beta / (s^n + \alpha_{n-1}s^{n-1} + \dots + \alpha_1s + \alpha_0) \quad (6.40)$$

where all the coefficients are real and positive, and with nonlinearity  $n(x)$ , chaotic motion was likely to exist if:

- (i) The Hurwitz sector of the linear system with  $n(x)$  replaced by  $K$  satisfies  $K_2 < K < K_1$ , where  $K_1 = \alpha_0 / -\beta$  and  $K_2$  is finite.
- (ii) The nonlinearity  $n(x)$  intersects the line through the origin of slope  $K_1$  at least once to ensure the existence of an equilibrium point.
- (iii) The slope of  $n(x)$  at all the intersection points with the lines of slope  $K_1$  and  $K_2$  should be either greater than the upper bound  $K_1$  or less than the lower bound  $K_2$ , thus ensuring the instability of the equilibria.

From their studies using the describing function Genesio and Tesi [6.21] concluded that the negative feedback system with nonlinearity  $n(x)$  and linear dynamics  $G(s)$  might be expected to exhibit chaos if

- (iv) The application of the DF method to the system gives a stable predicted limit cycle  $x^*(t)$ .
- (v) There exists an equilibrium point  $E$ , different from the one that generates the predicted limit cycle, which has one eigenvalue in the open right half plane and the others in the left half plane.
- (vi) The abscissa  $x_E$  belongs to the range of the predicted limit cycle, that is,  $x^*(t) = x_E$  for some time  $t$ .
- (vii) As the filtering properties of  $G(s)$  deteriorate the motion may be expected to change from a limit cycle to chaotic motion.

The investigations in [6.22] considered,  $n(x)$ , to be a relay and used the transfer function form of equation (6.40) for  $G(s)$ . It was conjectured that for  $n > 2$  the system was likely to possess a region of chaotic motion if

- (i) Equation (6.40) has at least one pair of complex conjugate roots with positive real part
- (ii) The Hurwitz sector lies within the nonlinearity sector

Based on these assumptions 3<sup>rd</sup> and 4<sup>th</sup> order relay systems with chaos were synthesized. For the third order system with an ideal relay with unit levels and  $\beta = 1$ , then the upper and lower bounds of the Hurwitz sector are  $K_1 = \alpha_0$  and  $K_2 = \alpha_0 - \alpha_1\alpha_2$ . Satisfaction of condition (iii) above requires  $0 < K_2 < K_1$ , which gives  $\alpha_0 > \alpha_1\alpha_2$ . Thus if the transfer function is chosen with  $\alpha_0 = 4$  and  $\alpha_1 = 1.25$  chaotic motion may exist for  $0 < \alpha_2 < 3.2$ .

The equilibria are stable for  $\alpha_2 > 3.2$ , become neutrally stable with roots at  $-3.2$  and  $\pm 1.118j$  for  $\alpha_2 = 3.2$ , and as  $\alpha_2$  decreases the complex roots have positive real parts. Using the program based on the Tsytkin method no limit cycle was found for  $\alpha_2 > 3.2$  but several unstable limit cycle solutions were found as  $\alpha_2$  was decreased below  $3.2$ . In particular an almost sinusoidal unstable limit cycle was found together with several 'spiral' type unstable limit cycles. When  $\alpha_2$  was reduced to  $2.8$  one of the 'spiral' limit cycles became stable and is shown in Figure 6.9. This may be referred to as a 3 spiral because of the number of oscillations. For  $2.8 < \alpha_2 < 3.2$  the various unstable limit cycles had different numbers of spirals apart from the almost sinusoidal one which bounded all the others. Simulations starting with initial conditions inside this bounding limit cycle resulted in chaotic motion with the motion 'jumping' between different 'spiral' type unstable limit cycles as shown in Figure 6.10.

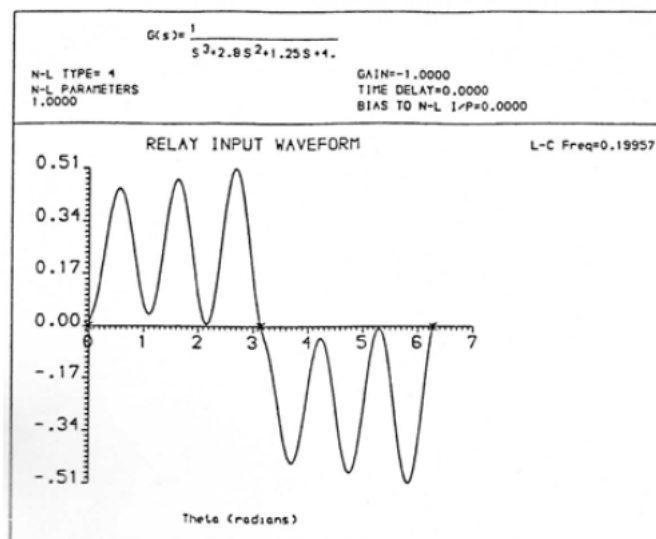


Figure 6.9 Three spiral limit cycle (taken from reference 6.19)

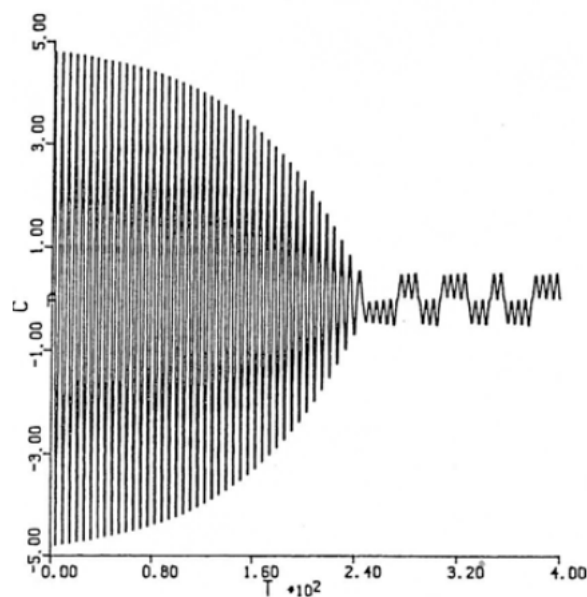


Figure 6.10 Chaotic spiral type motion (taken from reference 19)

### 6.6.6 Example 6 – Unstable Limit Cycles and Simulation

An interesting aspect of the above theory for finding limit cycles in relay systems is that both their waveform and stability can be accurately calculated. Stable limit cycles are normally not difficult to find in simulation studies as they may be reached from various initial conditions, unstable limit cycles are, however, much more difficult to locate. Since knowledge of the exact limit cycle solution allows initially conditions to be used in a simulation which lie on an unstable limit cycle, reference 6.23 studied the ability of a simulation to track the limit cycle. It was found that longer tracking was achieved, as expected, by using a smaller step size and also if the magnitude of the largest eigenvalue of the stability matrix  $Q$  was not significantly greater than unity.

## 6.7 Forced oscillations

The only modification required to be made to the method in order to investigate forced oscillations is that the relay input becomes a combination of the feedback signal and the forcing signal. Thus, if the relay input,  $x(t)$ , is the error signal of the feedback loop then  $x(t) = r(t) - c(t)$ . If a relay output waveform is assumed which switches when the input has a value of  $\Delta$  at the switching time  $\Delta t$  then  $r(\Delta t) - c(\Delta t) = \Delta$  and  $\dot{x}(t)$  must have the appropriate sign. An example is given in reference 6.12 for the theoretical solutions to applying a sinusoidal input to a feedback system having a relay with dead zone. The system gain is such that it has a limit cycle with no input. A sinusoidal input with a frequency of roughly three times that of the limit cycle is applied to the system and its amplitude slowly increased. Assuming initially a relay output with switchings appropriate to the limit cycle the plot of the relay input as the input amplitude increases contains more of the input frequency which eventually leads to a solution with false switchings. It is then assumed that the relay output will have two rather than one pulse per half cycle and a solution confirmed. Thus, embedding the theory in software with graphical output enables some results to be obtained for forced responses both for systems which may be stable or possess a limit cycle in the autonomous state.

## 6.8 Conclusions

This chapter has shown how an analysis based on a method originally formulated by Tsympkin can be used to find the exact form of limit cycles in relay systems. In addition to evaluating the limit cycles the method also allows their stability to be found. The evaluation of the limit cycles, however, requires investigation of the solution of nonlinear algebraic equations whose number increases with the number of pulses per period at the output of the relay. This is not a problem if the approach is imbedded in computer programs with interactive graphics as described above for some interesting examples studied using software written in Fortran. It is also possible to extend the method to cover systems with more than one relay as well as relay feedback systems with harmonic inputs.

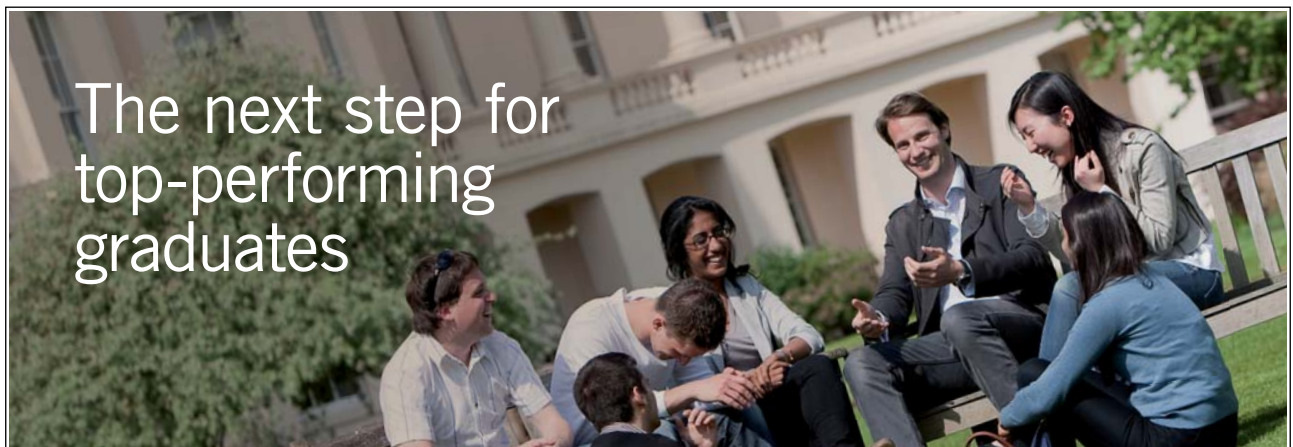
## 6.9 References

- 6.1 Bohn, EK 1963, *The transform Analysis of Linear Systems*, Addison Wesley, New York
- 6.2 Hamel, B 1950, *Etude Mathematique des Systemes a plusieurs Degrees de Liberte Decrits par des Equations Lineaires avec un Terme de Commande Discontinu*. Proc.Journees d'Etude des Vibrations, AERA,Paris.
- 6.3 Tsympkin, YaZ 1984, *Relay Control Systems* Cambridge University Press, Cambridge, England (English version of much earlier Russian edition).



- 6.4 Chung, JK-C & Atherton, DP 1966, The Determination of Periodic Modes in Relay Systems Using the State Space Approach, International Journal of Control, Vol 4, pp 105-126.
- 6.5 Majhi, S & Atherton, DP 2000, On-Line Tuning of Controllers for An Unstable FOPDT Process, IEE Proceedings Control Theory and Applications, Vol. 147, No.4, pp. 421-427.
- 6.6 Atherton, DP 1966, Conditions for Periodicity in Control Systems Containing Several Relays, Paper 28E, 3rd IFAC Congress, London,
- 6.7 Atherton, DP 1982, Oscillations in Relay Systems, Trans InstMC, Vol 3, pp 171-184.
- 6.8 Wadey, M 1984, Extensions of Tsytkin's Method for Computer Aided Control System Design, DPhil Thesis University of Sussex, England.
- 6.9 Willems, JL 1970, Stability Theory of Dynamical Systems. pp 112-113. Nelson, London
- 6.10 Balasubramanian, R 1981, Stability of limit cycles in feedback systems containing a relay IEE Proc. D Control Theory and Appl. Vol 128, pp24-29.
- 6.11 Wadey, MD & Atherton, DP 1982, Limit Cycles in Relay Systems, American Control Conference, Arlington, Virginia, USA.
- 6.12 Atherton, DP & Wadey, MD 1982, Computer Aided Analysis and Design of Relay Systems. IFAC Symposium on CAD of MV Tech. Systems, Purdue, USA, pp 355-360.
- 6.13 Atherton, DP 1993, Analysis and Design of Relay Control Systems, CAD for Control Systems, Chapter 15, Marcel Dekker, pp 367-394.
- 6.14 Rao, UM & Atherton, DP 1977, Stability Analysis of a Structurally Flexible Spacecraft. 20th Midwest Symposium on Circuits and Systems, Lubbock, Texas, pp 652-656.
- 6.15 Rao, UM and Atherton, DP, Multi-Pulse Oscillations in Relay Systems, Proceedings 7th IFAC World Congress, Helsinki 1978, Vol 3, Pergamon Press, pp 1747-1754.

Please click the advert



### Masters in Management



Designed for high-achieving graduates across all disciplines, London Business School's Masters in Management provides specific and tangible foundations for a successful career in business.

This 12-month, full-time programme is a business qualification with impact. In 2010, our MiM employment rate was 95% within 3 months of graduation\*; the majority of graduates choosing to work in consulting or financial services.

As well as a renowned qualification from a world-class business school, you also gain access to the School's network of more than 34,000 global alumni – a community that offers support and opportunities throughout your career.

For more information visit [www.london.edu/mm](http://www.london.edu/mm), email [mim@london.edu](mailto:mim@london.edu) or give us a call on **+44 (0)20 7000 7573**.

\* Figures taken from London Business School's Masters in Management 2010 employment report

- 6.16 Rao, UM, Atherton, DP & Balasubramanian, R 1978, Simulation and Estimation Studies of a Satellite Attitude Control System. IMACS Symposium on Simulation of Control Systems, Vienna, pp 215-221.
- 6.17 Moeini, A & Atherton, DP 1996, The Determination of Limit Cycles in Mode Dependent Relay Systems, IFAC Congress, San Francisco, Vol E, pp 43-48.
- 6.18 Atherton, DP, McNamara, OP, Wadey, MD & Goucem, A 1985, SUNS: the Sussex University Nonlinear Control Systems Software, 3rd IFAC Symposium on Computer Aided Design in Control Engineering Systems, Copenhagen, pp 173-178.
- 6.19 Atherton, DP 1990, Predicting the Autonomous Behaviour of Simple Nonlinear Feedback Systems, IMSE-90 Conference, Arlington, Texas. Published in App. Mech. Review, Vol 43, October 1990, pp 251-260.
- 6.20 Choudhury, SK & Atherton, DP 1974, Limit Cycles in High-Order Nonlinear Systems. Proceedings IEE, Vol 121, pp 717-724.
- 6.21 Genesio, R & Tesi, A 1992, Harmonic-Balance Methods for the Analysis of Chaotic Dynamics in Nonlinear-Systems, Automatica, Vol 28, pp531-548.
- 6.22 Amrani, D & Atherton, DP 1989, Designing Autonomous Relay Systems With Chaotic Motion, 28th IEEE Conference on Decision and Control, Florida, Vol 1, pp 512-517.
- 6.23 Wadey, MD & Atherton, DP 1986, A Simulation Study of Unstable Limit Cycles, IFAC/IMACS Symposium on Control Systems, Vienna, pp 213-218.

## 6.10 Appendix

The appendix presents formula from which the A loci of various transfer functions can be evaluated. This is done by putting any transfer function into partial fractions in terms of the simple transfer functions given in Table 6.1 and then using the results for their A loci given in Tables 6.2 and 6.3.

Type No.	Transfer Function	Type No.	Transfer Function
1	$1/sT$	6	$sT/(1 + sT)^2$
2	$1/(1 + sT)$	7	$1/s^3T^3$
3	$1/sT(1 + sT)$	8	$1/(1 + sT)^3$
4	$1/s^2T^2$	9	$1/(s^2 + 2\zeta s\omega_o + \omega_o^2)$
5	$1/(1 + sT)^2$	10	$s/(s^2 + 2\zeta s\omega_o + \omega_o^2)$

**Table 6.1** Basic Transfer Function Types.

The results for transfer functions with real poles are defined in terms of sine and cosine series and for the complex poles in terms of some parameters. The results are given for the A° loci in Table 6.2. The definitions of the sine and cosine series and the results for the transfer functions are given in Tables 6.2(a) and 6.2(b), where  $\lambda = \omega T$ . The parameters for evaluating the complex transfer functions are given in Table 6.2(c).

Similar results for the A loci are given in Tables 6.3, 6.3(a), 6.3(b) and 6.3(c), respectively.

<u>A° Loci</u>		
Type No.	Re A <sub>G</sub> <sup>°</sup> [θ, ω]	Im A <sub>G</sub> <sup>°</sup> [θ, ω]
1	$-\lambda^{-1}S_{-1,0}^{\circ}$	$-\lambda^{-1}C_{-2,0}^{\circ}$
2	$C_{0,1}^{\circ}-\lambda S_{1,1}^{\circ}$	$-S_{-1,1}^{\circ}-\lambda C_{0,1}^{\circ}$
3	$-\lambda^{-1}S_{-1,1}^{\circ}-C_{0,1}^{\circ}$	$-\lambda^{-1}C_{-2,0}^{\circ}+S_{-1,1}^{\circ}+C_{0,1}^{\circ}$
4	$-\lambda^{-2}C_{-2,0}^{\circ}$	$\lambda^{-2}S_{-3,0}^{\circ}$
5	$-2\lambda S_{1,2}^{\circ}+C_{0,2}^{\circ}-\lambda^2 C_{2,2}^{\circ}$	$-2\lambda C_{0,2}^{\circ}-S_{-1,2}^{\circ}+\lambda^2 S_{1,2}^{\circ}$
6	$C_{0,1}^{\circ}-\lambda S_{1,1}^{\circ}+2\lambda S_{1,2}^{\circ}$ $-C_{0,2}^{\circ}-\lambda^2 C_{2,2}^{\circ}$	$-S_{-1,1}^{\circ}-\lambda C_{0,1}^{\circ}+2\lambda C_{0,2}^{\circ}$ $+S_{-1,2}^{\circ}-\lambda^2 S_{1,2}^{\circ}$
7	$\lambda^{-3}S_{-3,0}^{\circ}$	$\lambda^{-3}C_{-4,0}^{\circ}$
8	$C_{0,2}^{\circ}-4\lambda^2 C_{2,3}^{\circ}-4\lambda S_{1,3}^{\circ}$ $+ \lambda S_{1,2}^{\circ}$	$-3\lambda C_{0,2}^{\circ}+4\lambda^3 C_{2,3}^{\circ}-S_{-1,2}^{\circ}$ $+4\lambda^2 S_{1,3}^{\circ}$
9	$\frac{\pi}{4b\omega} \left\{ \frac{c_1}{d} \right\}$	$\frac{\pi}{4\omega_0} \left[ \frac{b(c_2-d) - ac_1}{bd} \right]$
10	$-\frac{\pi}{4b\omega} \left[ \frac{ac_1 + bc_2}{d} \right]$	$\frac{\pi}{4b} \left\{ \frac{c_1}{d} \right\}$

Table 6.2 Results for A° loci

$$\begin{aligned}
S_{j,k}^{\circ} &= S_{j,k}^{\circ}(\theta, \lambda) = \sum_{n=1}^{\infty} \frac{n^j \sin n\theta}{(1+n^2\lambda^2)^k} \quad \text{for } 0 < \theta < \pi \\
S_{j,k}^{\circ}(-\theta, \lambda) &= -S_{j,k}^{\circ}(\theta, \lambda) \\
S_{-1,0}^{\circ}(\theta, \lambda) &= \pi/4 \\
S_{1,1}^{\circ}(\theta, \lambda) &= \pi \cosh[(\pi-2\theta)/2\lambda] / 4\lambda^2 \cosh(\pi/2\lambda) \\
S_{-1,1}^{\circ}(\theta, \lambda) &= S_{-1,0}^{\circ} - \lambda^2 S_{1,1}^{\circ} \\
S_{1,2}^{\circ}(\theta, \lambda) &= \{ \pi\theta \sinh[(\pi-2\theta)/2\lambda] \\
&\quad + \pi^2 \sinh(\theta/\lambda) / 2 \cosh(\pi/2\lambda) \} / 8\lambda^3 \\
&\quad \times \cosh(\pi/2\lambda) \\
S_{-1,2}^{\circ}(\theta, \lambda) &= S_{-1,1}^{\circ} - \lambda^2 S_{1,2}^{\circ} \\
S_{1,3}^{\circ}(\theta, \lambda) &= \{ \pi^2 \lambda \sinh(\theta/\lambda) + 2\pi\theta \lambda \cosh(\pi/2\lambda) \\
&\quad \times \sinh[(\pi-2\theta)/2\lambda] - 2\pi^2 \theta \cosh(\theta/\lambda) \\
&\quad + 2\pi\theta^2 \cosh(\pi/2\lambda) \cosh[(\pi-2\theta)/2\lambda] \\
&\quad + \pi^3 \sinh(\theta/\lambda) \tanh(\pi/2\lambda) \} / 64\lambda^4 \cosh^2(\pi/2\lambda) \\
S_{-3,0}^{\circ}(\theta, \lambda) &= (\pi^2 \theta - \pi \theta^2) / 8
\end{aligned}$$

Table 6.2(a) Results for Sine Series



$$C_{j,k}^{\circ} = C_{j,k}^{\circ}(\theta, \lambda) = \sum_{n=1}^{\infty} \frac{n^j \cos n\theta}{(1+n^2 \lambda^2)^k} \quad \text{for } 0 < \theta < \pi$$

$$C_{j,k}^{\circ}(-\theta, \lambda) = C_{j,k}^{\circ}(\theta, \lambda)$$

$$C_{0,1}^{\circ}(\theta, \lambda) = \{\pi \sinh[(\pi-2\theta)/2\lambda]\} / 4\lambda \cosh(\pi/2\lambda)$$

$$C_{-2,0}^{\circ}(\theta, \lambda) = (\pi^2 - 2\pi\theta) / 8$$

$$C_{2,2}^{\circ}(\theta, \lambda) = \{\pi\lambda \sinh[(\pi-2\theta)/2\lambda] - \pi\theta \cosh[(\pi-2\theta)/2\lambda] + [\pi^2 \cosh(\theta/\lambda)] / 2\cosh(\pi/2\lambda)\} / 8\lambda^4 \cosh(\pi/2\lambda)$$

$$C_{0,2}^{\circ}(\theta, \lambda) = C_{0,1}^{\circ} - \lambda^2 C_{2,2}^{\circ}$$

$$C_{2,3}^{\circ}(\theta, \lambda) = \{2\pi\lambda^2 \sinh[(\pi-2\theta)/2\lambda] \cosh(\pi/2\lambda) - \pi^2 \lambda \cosh(\theta/\lambda) + 2\pi\theta \lambda \cosh[(\pi-2\theta)/2\lambda] \cosh(\pi/2\lambda) - 2\pi^2 \theta \sinh(\theta/\lambda) - 2\pi\theta^2 \sinh[(\pi-2\theta)/2\lambda] \cosh(\pi/2\lambda) + \pi^3 \cosh(\theta/\lambda) \tanh(\pi/2\lambda)\} / 64\lambda^5 \cosh^2(\pi/2\lambda)$$

$$C_{-4,0}^{\circ}(\theta, \lambda) = (\pi^4 - 6\pi^2 \theta^2 + 4\pi\theta^3) / 96$$

$$C_{2,1}^{\circ}(\theta, \lambda) = \{-\pi \sinh[(\pi-2\theta)/2\lambda]\} / 4\lambda^3 \cosh(\pi/2\lambda).$$

Table 6.2(b) Results for Cosine Series

$$\begin{aligned}
 a &= \zeta \omega_o, & b &= \omega_o \sqrt{1 - \zeta^2} \\
 d &= \cosh(a\pi/\omega) + \cos(b\pi/\omega) \\
 c_1 &= \exp(a\theta/\omega) \{ \exp(-a\pi/\omega) \sin(b\theta/\omega) - \sin[(\pi-\theta)b/\omega] \} \\
 c_2 &= \exp(a\theta/\pi) \{ \exp(-a\pi/\omega) \cos(b\theta/\omega) + \cos[(\pi-\theta)b/\omega] \}
 \end{aligned}$$

Table 6.2(c) Parameters for Complex Pole Transfer Functions



You're full of *energy*  
and *ideas*. And that's  
just what we are looking for.

Looking for a career where your ideas could really make a difference? UBS's Graduate Programme and internships are a chance for you to experience for yourself what it's like to be part of a global team that rewards your input and believes in succeeding together.

Wherever you are in your academic career, make your future a part of ours by visiting [www.ubs.com/graduates](http://www.ubs.com/graduates).

[www.ubs.com/graduates](http://www.ubs.com/graduates)



© UBS 2010. All rights reserved.

Please click the advert

<u>A Loci</u>		
Type	Re $A_G[\theta, \omega]$	Im $A_G[\theta, \omega]$
1	$-\lambda^{-1} s_{-1,0}$	$-\lambda^{-1} c_{-2,0}$
2	$c_{0,1} - \lambda s_{1,1}$	$-s_{-1,1} - \lambda c_{0,1}$
3	$-\lambda^{-1} s_{-1,1} - c_{0,1}$	$-\lambda^{-1} c_{-2,0} + s_{-1,1} + \lambda c_{0,1}$
4	$-\lambda^{-2} c_{-2,0}$	$\lambda^{-2} s_{-3,0}$
5	$-2\lambda s_{1,2} + c_{0,2} - \lambda^2 c_{2,2}$	$-2\lambda c_{0,2} - s_{-1,2} + \lambda^2 s_{1,2}$
6	$c_{0,1} - \lambda s_{1,1} + 2\lambda s_{1,2}$ $- c_{0,2} - \lambda^2 c_{2,2}$	$-s_{-1,1} - \lambda c_{0,1} + 2\lambda c_{0,2}$ $+ s_{-1,2} - \lambda^2 s_{1,2}$
7	$\lambda^{-3} s_{-3,0}$	$\lambda^{-3} c_{-4,0}$
8	$c_{0,2} - 4\lambda^2 c_{2,3} - 4\lambda s_{1,3}$ $+ \lambda s_{1,2}$	$-3\lambda c_{0,2} + 4\lambda^3 c_{2,3} - s_{-1,2}$ $+ 4\lambda^2 s_{1,3}$
9	$-\left[ \frac{1}{2\omega_0^2} + \frac{\pi c_3 d_1}{2b\omega d_0} \right]$	$\frac{1}{\omega_0} \left[ -(\operatorname{sgn} \theta) \frac{\pi -  \theta }{2} + \frac{\omega a}{\omega_0^2} \right.$ $\left. + \frac{\pi [c_1 - c_2] d_1}{2b d_0} \right]$
10	$\frac{\pi [c_4 - c_5] d_1}{2b\omega d_0}$	$-\left[ \frac{\omega}{2\omega_0^2} + \frac{\pi c_3 d_1}{2b d_0} \right]$

Table 6.3 Results for A loci

$$\begin{aligned}
s_{j,k} &= s_{j,k}(\theta, \lambda) = \sum_{n=1}^{\infty} \frac{n^j \sin n\theta}{(1+n^2\lambda^2)^k} \quad \text{for } 0 < \theta < 2\pi \\
s_{j,k}(-\theta, \lambda) &= -s_{j,k}(\theta, \lambda) \\
s_{-1,0}(\theta, \lambda) &= (\pi - \theta)/2 \\
s_{1,1}(\theta, \lambda) &= (\pi/2\lambda)^2 \sinh[(\pi - \theta)/\lambda] / \sinh(\pi/\lambda) \\
s_{-1,1}(\theta, \lambda) &= s_{-1,0}(\theta, \lambda) - \lambda^2 s_{1,1}(\theta, \lambda) \\
s_{1,2}(\theta, \lambda) &= -(\pi/4\lambda^3) (\pi \sinh(\theta/\lambda) - \theta \sinh(\pi/\lambda) \cosh \\
&\quad \times [(\pi - \theta)/\lambda]) / \sinh^2(\pi/\lambda) \\
s_{-1,2}(\theta, \lambda) &= s_{-1,1}(\theta, \lambda) - \lambda^2 s_{1,2}(\theta, \lambda) \\
s_{1,3}(\theta, \lambda) &= -(\pi/16\lambda^4) ((\theta \sinh(\pi/\lambda) \cosh[(\pi - \theta)/\lambda] \\
&\quad - \pi \sinh(\theta/\lambda)) (\lambda \sinh(\pi/\lambda) + 2\pi \cosh(\pi/\lambda)) \\
&\quad + \sinh(\pi/\lambda) (\pi\theta \cosh(\theta/\lambda) \\
&\quad - \pi\theta \cosh[(2\pi - \theta)/\lambda] + \theta^2 \sinh(\pi/\lambda) \\
&\quad \times \sinh(\pi - \theta)/\lambda)) / \sinh^3(\pi/\lambda) \\
s_{-3,0}(\theta, \lambda) &= [(\theta - \pi)^3 - \pi^2\theta + \pi^3]/12
\end{aligned}$$

Table 6.3(a) Results for Sine Series



$$C_{j,k} = C_{j,k}^{\circ}(\theta, \lambda) = \sum_{n=1}^{\infty} \frac{n^j \cos n\theta}{(1+n^2 \lambda^2)^k} \quad \text{for } 0 \leq \theta < 2\pi$$

$$C_{j,k}(-\theta, \lambda) = C_{j,k}(\theta, \lambda)$$

$$C_{0,1}(\theta, \lambda) = (\pi/2\lambda) \cosh[(\pi-\theta)/\lambda] / \sinh(\pi/\lambda) - (1/2)$$

$$C_{-2,0}(\theta, \lambda) = (\theta-\pi)^2/4 - \pi^2/12$$

$$C_{2,2}(\theta, \lambda) = -(\pi/4\lambda^4) \{ \pi \cosh(\theta/\lambda) - [\lambda \sinh(\pi/\lambda) \cosh[(\pi-\theta)/\lambda] - \theta \sinh(\pi/\lambda) \sinh[(\pi-\theta)/\lambda]] \} \\ \div \sinh^2(\pi/\lambda)$$

$$C_{0,2}(\theta, \lambda) = C_{0,1}(\theta, \lambda) - \lambda^2 C_{2,2}(\theta, \lambda)$$

$$C_{2,3}(\theta, \lambda) = \frac{\partial}{\partial \theta} [S_{1,3}(\theta, \lambda)]$$

$$C_{-4,0}(\theta, \lambda) = \{ 2\pi^2(\theta-\pi)^2 - (\theta-\pi)^4 \} / 48 - (7\pi^4/720)$$

Table 6.3(b) Results for Cosine Series

$$\begin{aligned}
 a &= \zeta \omega_0, \quad b = \omega_0 \sqrt{1 - \zeta^2} \\
 d_0 &= \cosh 2\pi a/\omega - \cos 2\pi b/\omega \\
 d_1 &= \exp[(\operatorname{sgn} \theta) a|\theta|/\omega] \\
 c_1 &= \{a \sin[b(|\theta| - 2\pi)/\omega] - (\operatorname{sgn} \theta) b \cos[b(|\theta| - 2\pi)/\omega]\} \\
 c_2 &= \{a \sin b|\theta|/\omega - (\operatorname{sgn} \theta) b \cos b|\theta|/\omega\} \\
 &\quad \{\exp[-(\operatorname{sgn} \theta) 2\pi a/\omega]\} \\
 c_3 &= \{\sin[b(|\theta| - 2\pi)/\omega] - \sin b|\theta|/\omega\} \exp \\
 &\quad \{-(\operatorname{sgn} \theta) 2\pi a/\omega\} \\
 c_4 &= \{a \sin[b(|\theta| - 2\pi)/\omega] + (\operatorname{sgn} \theta) b \cos[b(|\theta| - 2\pi)/\omega]\} \\
 c_5 &= \{a \sin b|\theta|/\omega + (\operatorname{sgn} \theta) b \cos b|\theta|/\omega\} \exp \\
 &\quad \{-(\operatorname{sgn} \theta) 2\pi a/\omega\}
 \end{aligned}$$

Table 6.3(c) Parameters for Complex Pole Transfer Functions

# 7 Controller Tuning from Relay Produced Limit Cycles

## 7.1 Introduction

Control theory is typically taught from the starting point of having a mathematical model of a plant, most often linear, and a controller is then designed to provide a specified performance. Techniques usually regarded as classical control consider simple forms of controller, such as phase lead or PID and the problem then becomes, once the specific form of controller has been chosen, to try and find parameters for the controller so that the system performance meets the specifications. When working with a mathematical model of the plant many methods of classical control can be used to find the controller parameters. Dependent on the specifications there may be no solution, a unique solution, or many solutions to the problem.

The latter, for example, could be the case if the only specification was on the response to a step input and required an overshoot less than 10% and a relatively slow settling time. On the other hand if it was required to minimise the integral squared error for the step input then there would be a unique or, possibly no feasible, solution. The most common adjustable parameter controller is the PID, with the three terms P- proportional, I – integral and D – derivative. It has been used in the process industry since the 1940's first in pneumatic, then successively in vacuum tube, transistor, integrated circuit, and now usually in microprocessor form; an implementation that allows significantly more functionality to be included.

Please click the advert



Discover the truth at [www.deloitte.ca/careers](http://www.deloitte.ca/careers)

**Deloitte.**

© Deloitte & Touche LLP and affiliated entities.

Download free ebooks at [bookboon.com](http://bookboon.com)

When commissioning or examining control loops engineers or plant operators often found that with suggested or existing PID parameter settings the loop performance was not satisfactory. Thus the obvious question arose, what could one do 'on the job' to get better parameter settings. Enter Ziegler and Nichols [7.1] who in their 1948 paper suggested two, in principle, simple approaches:-

1. Perform an open loop step response on the plant. From this response estimate the parameters of a first order plus dead time (FOPDT) model. Then set the controller parameters using given formulae involving the estimated plant parameters
2. With the plant under closed loop control put the controller into the P mode. Adjust the gain until an oscillation results, measure its frequency and record the P gain. Then set the controller parameters using given formulae involving the oscillation frequency and the P gain.

They referred to their approaches as parameter tuning, which seems logical as what was done, possibly by a technician, was perform a system test and from the readings calculate the controller parameters according to the given formulae, quite a standard engineering approach. Today many papers refer to parameter tuning when the values are calculated starting from a mathematical model of the plant, which does not seem the correct terminology, as one never sees the term used for obtaining the parameters of other fixed controllers, such as lead or lag controllers, by this approach.

The Z-N methods consist of two parts, which is not often clearly stated, namely a testing 'identification' procedure and then the use of their tabulated formula to set the parameters. The 'identification' procedures in the tests are:-

1. An estimated FOPDT model from the step response. This was chosen because it was judged to be a reasonable model for many simple processes, although other models could be fitted to the response if felt more appropriate. It should be noted that in the early days of control the standard experimental identification tests were step and/or frequency responses.
2. An estimate of the 'so-called' critical point (or gain margin point) of the plant, namely the gain and frequency value,  $\omega_c$ , when the plant gives a phase change of  $-180^\circ$ . This, is of course, a 'small bit of information' about the plant, and the major contribution of Z-N was to recognise the value of this information.

Their suggested controller parameters based on the measurements from both tests were based on designs to give a relatively high overshoot, around 25%, in the closed loop step response.

Given the identification result of 1, namely an FOPDT plant model, then there are basically an infinite number of analytical approaches which can be used to calculate PID controller parameters. The point is that the approach used should meet specified performance criteria which are application dependent and may, of course involve considerations related to input step and disturbance responses, noise, nonlinearity and other factors not usually mentioned in simple approaches. If the specifications address a desired closed loop step response behaviour, as many do, then an open loop frequency domain approach is indirect and in general will require iteration to achieve the desired closed loop performance.

Both the Z-N identification methods have problems in practice. In the first test noise in the measurement affects the accuracy of getting the FOPDT model and it cannot, of course, be used for an open loop unstable plant. If used as intended, however, for plants in which one has reasonable confidence have essentially the assumed dynamic behaviour, and not any 'black box' model, then the identified parameters normally provide a satisfactory model for analytical design methods [7.2]. The second test is extremely difficult to implement in practice as many process plants have slow dynamics, so the reaction to a change in loop gain is slow. Further the principle is based on linear dynamics and actuators invariably have dead zones and indeed without saturation of some form dangerous amplitude oscillations can result from too large an increase in the P gain. It is, of course, a test which can be performed ideally in simulation using a modern digital simulation language, although it was not easy on an analogue computer simulation and can be, as mentioned, a 'nightmare' in practice.

Recognising these practical difficulties with the test, Astrom and Hagglund [7.3] suggested replacing the P gain control by an ideal relay, which became simple to implement with the move to microprocessor controllers in the 80's. The question then arises, having replaced the P control by a relay, as to how one uses the resulting limit cycle to obtain information about the plant. The earlier chapters provide the information for doing this.

## 7.2 Knowledge from the Limit Cycle

This can be obtained by basing the analysis either on use of the DF method or the exact method for limit cycles covered in the previous chapter. The original proposal [7.3], usually known as relay autotuning was based on the DF method and like the original Z-N work assumed the plant was unknown but typically having a transfer function form common for the process industries. Recently other approaches have been suggested where by assuming a specific model form for the plant its parameters are estimated from measurements on the limit cycle.

### 7.2.1 Relay Autotuning

For a negative feedback loop containing a plant with transfer function,  $G(s)$ , and relay, use of DF analysis gives

$$N(a)G(j\omega) = -1 \quad (7.1)$$

which can be written

$$|N(a)||G(j\omega)| = 1 \text{ and } \text{Arg}N(a) + \text{Arg}G(j\omega) = -180^\circ \quad (7.2)$$

For an on-off relay of height  $\pm h$  and hysteresis  $\pm \Delta$  then

$$N(a) = \frac{4h}{a\pi} e^{-j\theta} \text{ where } \theta = \sin^{-1}(\Delta/a) \quad (7.3)$$

Thus when the nonlinearity is an ideal relay and denoting the limit cycle frequency by  $\omega_o$ , one has

$$\text{Arg}G(j\omega_o) = -180^\circ \text{ and } |G(j\omega_o)| = a\pi/4h \quad (7.4)$$

This shows that the DF analysis yields

$$\omega_c = \omega_o \text{ and } |G(j\omega_c)| = a\pi/4h = 1/K_c \quad (7.5)$$

for the frequency and magnitude of the critical point, where  $K_c$  is the critical gain. Due to the approximation of the DF method this result is approximate and is normally more accurate the nearer the limit cycle is to being sinusoidal.

It should also be noted that  $a$  is not the amplitude of the limit cycle at the relay input but the amplitude of its fundamental. Most practical implementations measure the peak to peak limit cycle amplitude and take  $a$  to be one half of this. It is possible to include additional computation software in the controller to obtain a better estimate of the fundamental  $a$  but in view of the inaccuracies in measurement it is probably not justifiable in many instances.. On the other hand sometimes easy corrections are possible [7.4] for example if the limit cycle is almost triangular, as it is for an FOPDT plant with small time delay, then one can calculate the fundamental from the peak value. Also the theory gives a better percentage accuracy for the frequency than the amplitude [7.5].

If time is no object in performing limit cycle experiments then one can:-

- (1) normally get more accurate measurements by averaging over several cycles.
- (2) vary the hysteresis in the relay to get additional results
- (3) include known extra linear elements in the loop to get additional results
- (4) get improved results from the DF method by making the limit cycle nearer to a sinusoid with the inclusion of a known tuned filter.

Please click the advert

It's only an opportunity if you act on it

IKEA.SE/STUDENT

© Inter IKEA Systems B.V. 2009



Typically (2) and (3) can be used to get additional points [7.6] to that of the critical point on the plant frequency response and iterative tuning of a resonant filter [7.6] can achieve (4) by producing a pure sinusoidal limit cycle at the plant critical frequency, which then yields the exact value of the critical point apart from, of course, measurement errors. The tuned filter was introduced in [7.6] in a paper concerned with the design of nonlinear PID controllers for nonlinear plants, where to get better information on the plant frequency response a sinusoidal input was required. A big advantage of the relay autotuning method is that the amplitude of the limit cycle can be controlled by the setting of the relay output levels, and this avoids one of the practical problems of the second method of Z-N. Further one can use this feature to estimate whether the plant behaves reasonably linearly and if not use the results to design nonlinear PID terms as discussed in section 9.5.

## 7.2.2 Plant Parameter Identification

It has been shown how the limit cycle in a loop with a known plant transfer function under relay control can be calculated exactly using the Tsytkin method or estimated using the DF method. In particular for the case of an odd symmetrical limit cycle the first method allows exact determination of the limit cycle amplitude and frequency at the relay input in terms of the plant parameters. Thus, the method can also be used for plant transfer function identification since measurement of the amplitude and frequency of the limit cycle will allow two unknown plant parameters to be found from the equations, for example the time delay and time constant in an FOPDT model when the plant d.c. gain is known. Analysis for an asymmetrical limit cycle and measurement of more parameters in the resulting limit cycle enables additional plant parameters to be calculated, for example four parameters in an SOPDT model.

Several papers [7.7, 7.8] describe the identification of plant parameters based on an assumed plant transfer function form, using exact limit cycle analysis and limit cycle measurements. From a practical viewpoint they have the following possible disadvantages.

- (1) The microprocessor controller software, as well as being able to 'measure' various limit cycle parameters, must be able to solve the appropriate nonlinear algebraic equations to get the plant parameters. Further it must contain a design procedure for calculating the controller parameters for the then known plant.
- (2) There will be errors in the measured limit cycle parameters, due to noise etc, and these may cause problems in solving the nonlinear algebraic equations, too large errors for example can result in no or physically unallowable solutions.

It is also possible to assume a plant transfer function and determine expressions for some limit cycle parameters using DF analysis rather than the exact approach. Again one can then use the equations 'in reverse' and calculate plant parameters from limit cycle measurements. If this is done then due to the approximation of the DF method, particularly when dealing with asymmetrical limit cycles, the problem cited in (2) is normally worse.

### 7.3 Tuning the Controller

Many design methods may be used to obtain controller parameters to meet given system specifications when the plant transfer function has been estimated by one of the methods in section 7.2.2. Both ‘classical’ and ‘modern methods’ will often turn out to require an iterative approach as they will invariably not address the required specifications directly but by inference. For example, if the design specification is with respect to the input step response, then if one uses a frequency domain design one will be relying on an imprecise link, say between phase margin and overshoot in the step response, or if one uses pole placement the effect of any closed loop zeros or neglected poles on the expected response will have to be checked.

The actual system behaviour will differ from the theoretical calculations due to the modelling errors of the identification and this, assuming the plant behaves reasonably linearly, can be taken into account by considering parameter uncertainty. Much is made about the concepts of robust control for such situations but no results address the real design problem, namely if the plant parameters vary over a certain range will the specifications, which normally relate to responses, not stability, still be met. Fortunately today Matlab is so fast that the robustness of a controller design can easily be checked simply by doing a set of Monte Carlo simulations over the expected range of variations in a few parameters. In fact a good designer may be able to identify the ‘worst case’, so that the design just meets the specification for this case and is better for other parameter values.

On the other hand if one uses the limit cycle test to estimate the critical point of the plant transfer function, two issues arise namely how accurate is the critical point estimate and secondly how, knowing just this bit of information, can the controller parameters be found so that the system meets the desired specifications. The first point has largely been addressed in the previous sections and to some extent how errors in the critical point value may affect the subsequent design can be examined by considering in simulation how the performance differs when selecting a few slightly different values for the critical point.

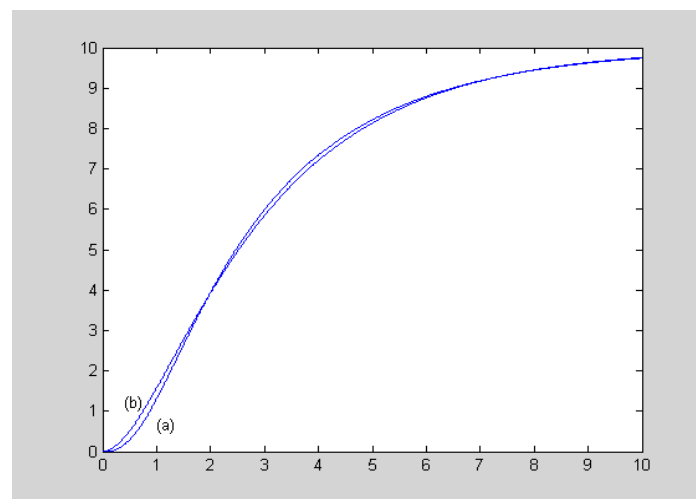
The second point is much more difficult, since even if a simple step response specification is given, one cannot assert for given PID parameters that if a plant with a given critical point has a closed loop step response with a given percentage overshoot then all plants with the same critical point will have the same overshoot. Certainly the closeness of the open loop frequency response of the compensated system to the  $(-1, j0)$  point has a significant effect on the form of the closed loop step response but the shape of the response in this region can differ significantly for different plant model transfer functions with the same critical point and the same compensator. It is therefore desirable in using critical point information to have some idea of the form of the plant transfer function. Having obtained the critical point all that can be done in a controller design algorithm is to select the controller parameters so that, subject to realisability constraints, the compensated frequency response has the critical frequency at a desired point in the Nyquist plane.

This point can be moved around, within a finite region, by adjusting the controller parameters. For the ideal PID controller  $K[1 + sT_i + (1/sT_d)]$ , Z-N chose to take  $T_i = 4T_d$  and it is easily shown that their choice for  $T_d$  and  $K$  places the critical frequency on the compensated locus at a gain of 0.66 and phase angle of  $-155^\circ$ . The corresponding point for their suggested PI parameters is 0.46 at a phase angle of  $-191^\circ$ . These choices, as for their first method, tend to produce closed loop step responses with a relatively high overshoot.

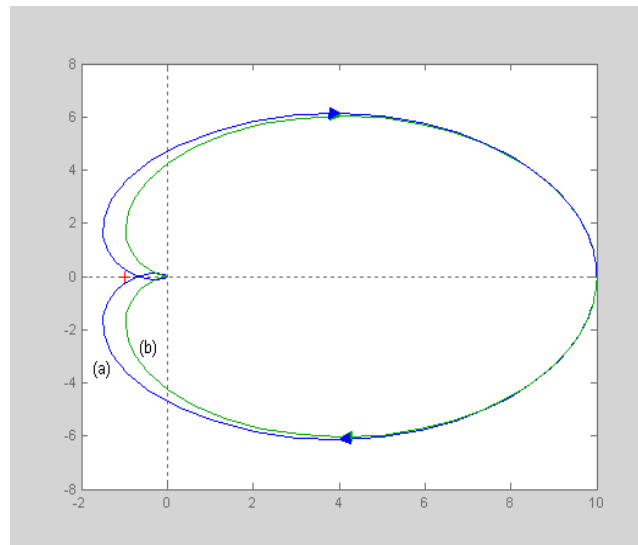


Thus the success of PID controller design based on critical point information is one of knowing where the critical frequency should be on the compensated locus to, hopefully, meet the design specifications. Given some idea of the plant transfer function it is relatively straightforward to come up with good strategies. For example, in reference 7.9 optimum parameters are found for PID controllers to minimise integral of error performance indices for a step input to a closed loop with an FOPDT plant. These are given for minimization of the ISE, ISTE and  $IST^2E$ . Typically as the time weighting increases the overshoot decreases, the rise time is slower and the settling time increases slightly. The location of the critical frequency on the compensated Nyquist locus has also been calculated for these designs so that critical frequency shifting designs can be used for other similar plant transfer functions. Minimization of the ITSE typically leads to a design with around 10% overshoot. Many different design approaches are possible for deciding where to place the critical frequency. Also, depending on the application different requirement specifications may have to be met.

One might be tempted to say that knowing a bit of information about a plant transfer function, namely its critical point, cannot be anything like as good as having an experimentally measured open loop plant step response, in principle a multiplicity of points. The problem is that measured step response results are not exact with large amounts of noise often present in industrial measured data. To illustrate a point consider the two step responses in Figure 7.1, curve (a) is for a chosen transfer function and curve (b) for a transfer function derived from it by a model reduction technique. On this basis, particularly if the response of (a) had resulted from measurements contaminated by noise, one would have clearly concluded that the transfer function of curve (b) would be a good model to use for controller design for the original transfer function (a). Now go to Figure 7.2 and look at the corresponding frequency responses. It can be seen that our previous conclusion is clearly not valid as the transfer function with the model (a) has a finite and model (b) an infinite gain margin. Thus a poor design for say, a constant gain controller, would result from using model (b), one in fact which would probably result in an unstable system with the chosen transfer function. The two transfer functions are  $\frac{13s + 10}{(s + 1)(0.5s + 1)(s^2 + 3s + 1)}$  and  $\frac{5.435}{s^2 + 1.739s + 0.5435}$ , respectively.



**Figure 7.1** Step responses of original and reduced models.



**Figure 7.2** Frequency responses of original and reduced models

Please click the advert

## YOUR CHANCE TO CHANGE THE WORLD

Here at Ericsson we have a deep rooted belief that the innovations we make on a daily basis can have a profound effect on making the world a better place for people, business and society. Join us.

In Germany we are especially looking for graduates as Integration Engineers for

- Radio Access and IP Networks
- IMS and IPTV

We are looking forward to getting your application!  
To apply and for all current job openings please visit our web page: [www.ericsson.com/careers](http://www.ericsson.com/careers)

ericsson.  
com

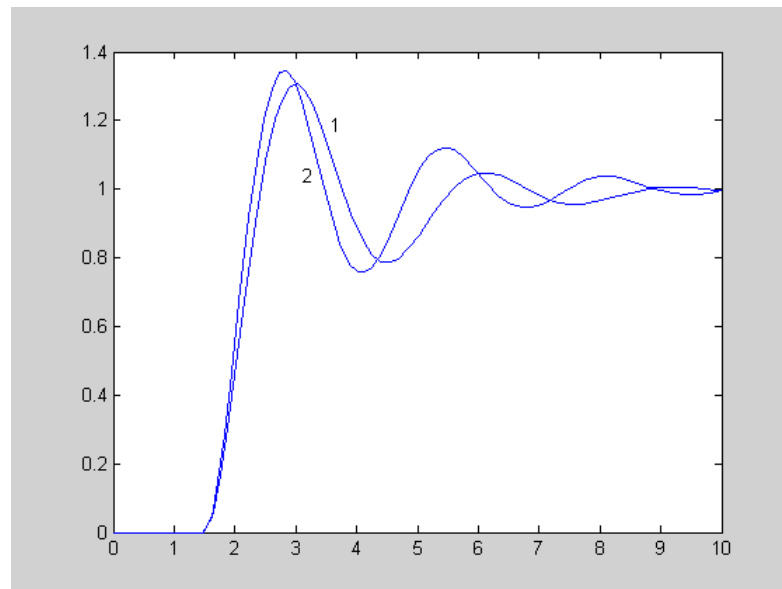


Download free ebooks at [bookboon.com](http://bookboon.com)

### 7.3.1 An Example

Consider for example a plant with the transfer function  $G(s) = \frac{K_p e^{-0.5s}}{(s+1)^2}$  with  $K_p = 1$ .

The objective is to show that good tuning can be obtained for a PID controller based on an autotuning test. Using a relay of unit height the measured values of the amplitude of the limit cycle,  $A$ , and frequency,  $\omega_0$ , from a simulation with no noise introduced were  $A = 0.287$  and  $\omega_0 = 1.90$ . Using equation (7.5) with  $a = A$  gives  $\omega_c = 1.90$  and  $K_c = 4.44$  which compare with the actual values of 1.92 rads/s and 4.69 for the known plant transfer function. The normalized gain  $\kappa = K_p K_c$  and, assuming  $K_p = 1$  has been found by other measurements,  $\kappa = 4.44$ . The tuning formula, suggested in reference 7.8, which gives optimum parameters for ISTE tuning (minimization of the integral squared of time multiplied by error) of a FOPDT plant based on  $\kappa, K_c$  and  $\omega_c$ , gives  $K = 2.55, T_i = 2.61$  and  $T_d = 0.40$ . The unit step set point response for these tuning parameters in the controller is shown in Figure 7.3 (response (1)) together with that for controller parameters of  $K = 2.42, T_i = 1.99$  and  $T_d = 0.63$  obtained by minimization of the ISTE criterion (response (2)) for this known plant transfer function. The results are very similar considering the approximations involved and the simplicity of the autotuning approach.



**Figure 7.3** Step response comparison for autotuning (1) and ISTE minimization (2)

## 7.4 Autotuning using the Relay in Parallel with the Controller

The standard autotuning approach is to replace the controller by a relay so that the forward path of the feedback loop contains a relay and the plant. Further if one wishes to check certain aspects of the compensated system, for example find its gain margin, then this can be done by placing the relay in series with the controller and plant [7.10]. These approaches involve what might be called separate tuning experiments but in some cases it may be appropriate to try and tune the system whilst it is still operating normally. This can be done by placing the relay in parallel with the controller [7.11]. This procedure was first suggested in reference 7.12, although in the paper the approach was used for plant parameter identification rather than critical point tuning.

When the relay is placed in parallel with the controller the feedback loop is as shown in Figure 7.4. To determine a limit cycle by DF analysis one simply opens the loop at the relay and calculates the transfer function,  $-G_e$ , between  $e_o$ , the relay output, and  $e_i$ , the relay input, in the figure.

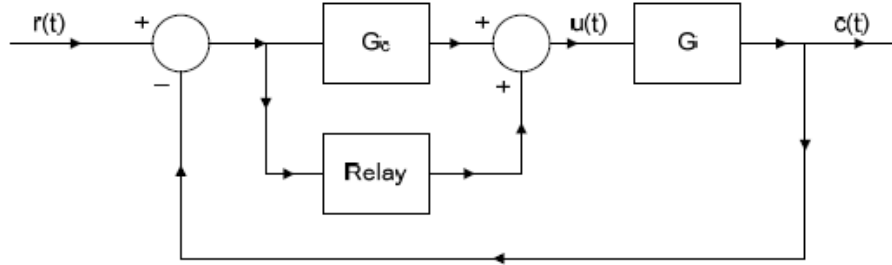


Figure 7.4 Relay in Parallel with the Controller

The basic equations (7.1-7.2) still apply but with  $G$  replaced by  $G_e$ . It is easily shown that

$$G_e = \frac{G}{1 + G G_c} \quad (7.6)$$

which can be written

$$G_e = \frac{1}{G_c} \times \frac{G G_c}{1 + G G_c} = \frac{1}{G_c} \times T_{cl} \quad (7.7)$$

where  $T_{cl}$  is the closed loop transfer function. Now consider the Nyquist diagram of  $G G_c$ , and denote its magnitude at the frequency of the limit cycle,  $\omega_o$ , by  $M$  and that of  $1 + G G_c$ , by  $\rho M$ , as shown in Figure 7.5. The magnitude equation in (7.2) can now be written

$$\frac{1}{|G_c|} \frac{M}{\rho M} = \frac{a\pi}{4h} \text{ which gives } \rho = \frac{4h}{a\pi |G_c|} \quad (7.8)$$

and the corresponding phase equation is

$$\phi = 180^\circ + \text{Arg}(1/G_c) \quad (7.9)$$

where  $-\phi$  is the closed loop phase angle shown in Figure 7.5.

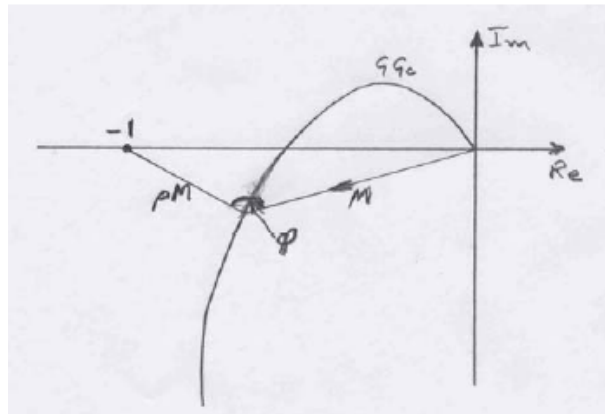


Figure 7.5 Nyquist diagram and triangle for calculations.

Since the transfer function in the controller is known  $|G_c|$  and  $\text{Arg } G_c$  are known at the frequency,  $\omega_o$ , so that equations (7.8) and (7.9) can be solved, respectively, for  $\rho$ , as  $a$  is known from the limit cycle measurement, and  $\phi$ . Using the cosine rule with the angle  $\phi$  for the triangle, since its sides are 1,  $M$  and  $\rho M$ , the value of  $M$  can be found.  $\text{Arg}(GG_c)$  can also be found from the triangle so that both the magnitude and phase of the frequency response of  $G$  at the frequency,  $\omega_o$ , can be calculated. This will only be the critical frequency for  $G$  if  $\text{Arg } G_c$  is zero, that is if the test can be done with the controller in the proportional mode. However, if  $\omega_o$  is near to the critical frequency then settings based on the critical frequency should produce satisfactory results. Alternatively, other theories can be produced for finding controller parameters with various types of plant for other frequencies on the plant frequency response locus.

Please click the advert

SIMPLY CLEVER

ŠKODA



We will turn your CV into  
an opportunity of a lifetime

Do you like cars? Would you like to be a part of a successful brand?  
We will appreciate and reward both your enthusiasm and talent.  
Send us your CV. You will be surprised where it can take you.

Send us your CV on  
[www.employerforlife.com](http://www.employerforlife.com)



### 7.4.1 An Example

As an example results from simulations of the FOPDT plant with transfer function  $G(s) = \frac{e^{-0.5s}}{s+1}$  and a relay with  $h = 1$  in parallel with the PID controller having  $T_i = 4T_d$ , so that  $G_c(s) = \frac{K(1 + 2sT_d)^2}{4sT_d}$ , are given.

The critical frequency and gain of this plant are, respectively, 3.67 rad/s and 3.81 and for the simulation the parameters of the controller were set at the Z-N values of  $K = 2.284$  and  $T_d = 0.214$ . For the input step response this results in an overshoot of 52.5%. The amplitude and frequency of the limit cycle at the input to the controller, which is shown in Figure 7.6, were measured as 0.680 and 4.30 rad/s.

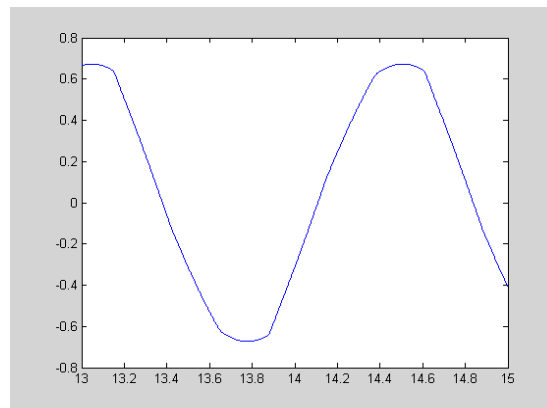
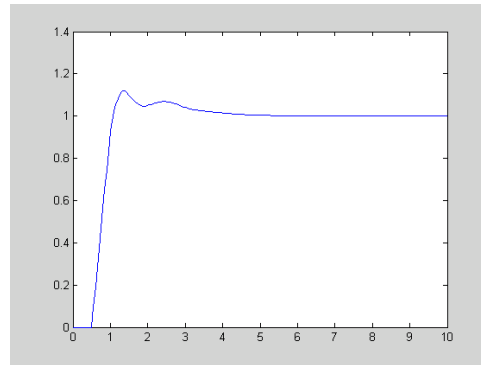


Figure 7.6 Plot of limit cycle

For the controller at  $\omega = 4.30$ ,  $|G_c(j\omega)| = 2.72$  and  $\text{Arg}G_c(j\omega) = 33.0^\circ$  giving  $\rho = 0.688$  and  $\varphi = 147^\circ$  from eqns. (7.8) and (7.9), respectively. Using the cosine formula for the triangle of Figure 7.5 gives  $M$  the magnitude of  $GG_c$  at  $\omega = 3.40$  as 0.617, and applying the formula again gives  $\alpha$ , the angle below the real axis of  $M$  equal to  $13.4^\circ$ , so that  $\text{Arg}GG_c = -166.6^\circ$ . Using the known values for the magnitude and argument of  $G_c$ , then gives the magnitude and argument of  $G$  as 0.227 and  $-199.6^\circ$ . Despite the fact that the waveform is not a good sinusoid these values are within 1% of the calculated values from the transfer function. The problem now becomes how should this information be used to set the controller parameters?

Here use is made of the ISTE critical point design procedure of reference 7.8. For this plant the critical point needs to be moved to 0.58 at  $-153^\circ$ , which means the PID gives a phase lead of  $27^\circ$ . Thus for a higher frequency it will give more phase lead, so that a reasonable suggestion for the frequency 3.40 is to increase the shift by around 25% to  $35^\circ$ . This gives a phase of approximately  $-165^\circ$  ( $199.6^\circ - 35^\circ$ ) for this frequency on the compensated frequency response locus. Since this point is nearer  $(-1, 0)$  than the critical frequency point by  $12^\circ$  a slight reduction in the magnitude by around 10% to the value of 0.53 is taken. Using the same  $T_i/T_d$  value as for the critical point shift design of reference 7.8, the required PID parameters are  $K = 1.94$ ,  $T_d = 0.205$  and  $T_i = 1.152$ , respectively. The input step response for this design is shown in Figure 7.7 and as expected is compatible with an ISTE design.



**Figure 7.7** Closed loop step response

## 7.5 Conclusions

This chapter has shown how information can be obtained about the linear dynamics of a feedback loop from measurements on limit cycles produced by adding a relay. In particular, the focus has been placed on using the results to set the parameters of a PID controller, however the concept can be used for deciding on appropriate parameters for other fixed structure controllers. With modern microprocessor controllers the approach is relatively easy and provides a good method for the automatic tuning of controller parameters without the need for mathematical modelling of the plant. Care, however, has to be taken to ensure that safe plant operation will be possible under relay control.

## 7.6 References

- 7.1 Ziegler, JG & Nichols, NB 1942, Optimum Settings for Automatic Controllers, Trans ASME, Vol 64, pp 759-768.
- 7.2 Atherton, DP, 2009 Control Engineering Bookboon publications at [www.bookboon.com](http://www.bookboon.com)
- 7.3 Åström, KJ & Hägglund, T 1984, Automatic Tuning of Simple Regulators with Specifications on Phase and Amplitude Margins, Automatica, Vol 20, pp 645-651.
- 7.4 Atherton, DP 1997, Improving the Accuracy of Autotuning Parameter Estimates. Proceedings, IEEE International Conference on Control Applications, Hartford, pp 51-56.
- 7.5 Popov, EP & Pal'tov, NP 1963, Approximate Methods for Analyzing Nonlinear Automatic System, English translation, Wright Patterson Air Force Base translation services.
- 7.6 Atherton, DP, Benouarets, M & Nanka-Bruce, O 1993, Design of Nonlinear PID Controllers for Nonlinear Plants, Proceedings IFAC World Congress '93, Sydney, Vol 3, pp 355-358.
- 7.7 Majhi, S & Atherton, DP 1999, Autotuning and Controller Design for Processes with Small Time Delays. Proc IEE Control Theory and Applications, Vol 146, pp 415-425.
- 7.8 Kaya, I & Atherton, DP 2001, Parameter Estimation from Relay Autotuning with Asymmetric Limit Cycle Data, J. of Process Control, Vol 11, pp 429-439.
- 7.9 Zhuang, M & Atherton, DP 1993, Automatic Tuning of Optimum PID Controllers. Proc.IEE Control Theory and Applications, Vol 140, pp 216-224.
- 7.10 Astrom, KJ 1996, Tuning and Adaptation, Plenary lecture, IFAC Congress, San Francisco.
- 7.11 Atherton, DP 2006, Relay Autotuning: An Overview and Alternative Approach, Ind. Eng. Chem. Res. Vol 45, pp. 4075-4080.
- 7.12 Majhi, S & Litz, L 2004, Relay Based Closed-loop Tuning of PID Controllers, Automatisierungstechnik, Vol.52, pp.202-20.

# 8 Absolute Stability Results

## 8.1 Introduction

It has been seen in previous chapters that the behaviour of the nonlinear feedback loop of Figure 3.1 and indeed, its stability, may be affected by the input. This has therefore led to two definitions of stability for a nonlinear system, the first stability of the autonomous system, that is with  $r(t) = 0$  in Figure 3.1, which is often referred to as stability of equilibrium and to which most early work was directed. The second form of stability is known as bounded-input bounded-output (BIBO) stability and requires the output to be bounded for a bounded-input signal. These two forms of stability are, of course, the same for a linear system; for which necessary and sufficient conditions for stability have been known for many years.

On the other hand the general nonlinear problem for the stability of equilibrium remains unsolved. Indeed until the development of frequency response based criteria in the 1960's [8.1] the only available results were those from the work of Lyapunov which were not easily usable by an engineer. The frequency response based criteria, which will be presented, only give sufficient conditions for stability and thus often produce very conservative results for a specific nonlinear characteristic as they only use limited information on the nonlinearity. On the other hand a mathematician may argue that the results are very robust because they ensure stability for a large number of nonlinearities which satisfy certain conditions. This, of course, means that the system can be proven to remain stable if the conditions are not violated by any perturbations to the nonlinearity.

The primary concern of this chapter is to cover frequency response absolute stability results as they allow comparisons to be made with both the DF method and limit cycle results for relay systems. In the next section Lyapunov's method is briefly covered as some of the original frequency domain results were obtained using it whilst others were obtained using functional analysis.

## 8.2 Lyapunov's Method

The direct or second method of Lyapunov for determining the stability of an autonomous system is a time domain method using a state space description. The problem is to determine the stability of an equilibrium state (singular point), which without loss of generality may be assumed to be the origin of the state space, for the system

$$\dot{x} = f(x) \tag{8.1}$$

Before considering Lyapunov's method some preliminary topics must be considered.



## 8.2.1 Definitions of Stability

**Definition 1**

The equilibrium state 0, or the equilibrium solution  $x(t) = 0$ , is called stable if for any given positive  $\varepsilon$ , there exists a positive  $\delta$ , such that  $\|x(t_0)\| < \delta$  implies  $\|x(t)\| < \varepsilon$  for all  $t > t_0$ , where  $\|\cdot\|$  denotes the Euclidian norm. It should be noted that a system exhibiting a limit cycle or solutions for which  $\|x(t)\|$  increases temporarily will be considered stable by this definition, when  $\varepsilon$  and  $\delta$  are appropriately chosen, but it rules out solutions which grow without bound. For a nonlinear system stability may only exist in some domain  $\|x\| < R$  of the state space.

**Definition 2**

The equilibrium state 0 is called asymptotically stable if it is stable and  $x(t) \rightarrow 0$  as  $t \rightarrow \infty$ .

**Definition 3**

The equilibrium state 0 is called asymptotically stable in the large, or globally asymptotically stable, if it is asymptotically stable and the domain  $R$  covers the whole state space.

Please click the advert

With us you can  
shape the future.  
Every single day.

For more information go to:  
[www.eon-career.com](http://www.eon-career.com)

Your energy shapes the future.

**e-on**

### 8.2.2 Positive Definite Functions

Let  $V(x)$  be a time invariant scalar function of the vector  $x$  and  $S$  be a closed bounded region in the  $x$  space containing the origin.

#### Definition 4

The function  $V(x)$  is positive definite (semi-definite) in  $S$  if, for all  $x$  in  $S$

- (i)  $V(x)$  has continuous partial derivatives with respect to the components of the vector  $x$
- (ii)  $V(0) = 0$
- (iii)  $V(x) > 0$  for  $x \neq 0$  (or  $V(x) \geq 0$ ).

If the inequality signs in (iii) are reversed then the conditions for a negative definite (or semi-definite) function are obtained. Thus, for example, in two space if  $V(x) = x_1^2 + x_2^2$  then it is positive definite as it is only zero at the origin, but the function  $V(x) = (x_1 + x_2)^2$  is only positive semi-definite as it is zero everywhere along the line  $x_1 + x_2 = 0$ .

A common quadratic form scalar function is

$$Q(x) = x^T Q x \quad (8.2)$$

where the matrix  $Q$ , which is  $n \times n$  for  $x$  an  $n$  vector, is chosen to be symmetric. In this case it is known that a necessary and sufficient condition for the quadratic form of (8.2) to be positive definite is that the determinants of all the principle leading minors of  $Q$  be positive. For it to be positive semi-definite all the principle minors, not just the leading ones, must be non-negative.

### 8.2.3 Lyapunov's Theorems

#### Theorem 1

The null solution, or equilibrium state of the origin, of the system of equation (8.1) is (asymptotically) stable if in some neighbourhood of the origin there exists a positive definite function  $V(x)$  such that its derivative  $\dot{V}(x)$  with respect to the solutions of equation (8.1) is negative (definite) semi-definite in that region. The conditions for asymptotic stability may be relaxed slightly by a theorem due to LaSalle which is given next.

#### Theorem 2

The null solution of the system of equation (8.1) is asymptotically stable if in some neighbourhood  $R$  of the origin there exists a positive definite function  $V(x)$  such that its derivative  $\dot{V}(x)$  is negative semidefinite in  $R$  and is non-zero in  $R$  along any solution of equation (8.1) other than the null solution. If the conditions of Theorem 2 are satisfied for all  $x$  then the equilibrium state is asymptotically stable in the large. The function  $V(x)$ , which satisfies the stable condition of Theorem 1, for the equation (8.1) is known as a Lyapunov function for the nonlinear equation. It should be noted that finding a Lyapunov function provides a sufficient condition for stability of equation (8.1). Thus the difficulty of applying Lyapunov's method is one of finding a 'good'  $V(x)$ . Failure to find one does not mean the system is unstable and finding one may result in a very conservative condition for stability.

### 8.3 Application of Lyapunov's Method

As might be expected it is possible to prove all the known results for the stability of a linear system using Lyapunov's method but good results for nonlinear systems, despite intensive research by mathematicians, are not easy to obtain.

#### 8.3.1 Linear System

For the linear system

$$\dot{x} = Ax \quad (8.3)$$

one can choose  $V(x) = x^T Q x$ . Differentiating and substituting for  $\dot{x}$  and  $\dot{x}^T$  gives

$$\dot{V}(x) = x^T (A^T Q + Q A) x = x^T P x \quad (8.4)$$

where

$$P = A^T Q + Q A \quad (8.5)$$

which is known as the Lyapunov equation, and which can be solved for  $Q$  given  $A$  and  $P$ .

A Lyapunov function for a linear system can be found by taking  $P = -I$ . For example if  $A = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix}$  then with  $P = -I$  the solution for  $Q$  is  $Q = \begin{pmatrix} 5/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}$  and

$$V(x) = (1/4)(5x_1^2 + 2x_1x_2 + x_2^2) \quad (8.6)$$

which is positive definite.

On the other hand if we had assumed any quadratic form for  $V(x)$  it may not have been possible to prove stability. For instance for

$$V(x) = x_1^2 + x_2^2 \quad (8.7)$$

Then

$\dot{V}(x) = 2x_1\dot{x}_1 + 2x_2\dot{x}_2 = 2x_1x_2 + 2x_2(-2x_1 - 3x_2) = -(2x_1x_2 + 6x_2^2)$  which is not negative definite and no conclusion can be drawn about stability. A physical interpretation can be found for this if the equations (8.6) and (8.7) are drawn on a phase plane for different constant values of  $V$  together with trajectories for the system. It will be seen that all the trajectories for the system cut the former constant  $V$  curves only once in the direction of decreasing  $V$ , which is not the case with the latter.

### 8.3.2 Nonlinear System

There have been many efforts to try and find suitable  $V$  functions for nonlinear systems. One of the most successful was due to Lur'e who for a feedback system with a single nonlinearity in the feedback path given by

$$\begin{aligned}\dot{x} &= Ax + bu \\ u &= n(z) \\ z &= c^T x\end{aligned}\tag{8.8}$$

used a  $V$  function of the form

$$V(x, z) = x^T Qx + \beta \int_0^z n(z) dz\tag{8.9}$$

He was able to obtain necessary conditions for  $V(x, z)$  to be a Lyapunov function with (a) the nonlinearity confined to the sector  $[0, K]$  and all the eigenvalues of  $A$  in the left half plane and (b) the nonlinearity confined to the sector  $[\epsilon, K]$  and all the eigenvalues of  $A$  confined to the left half plane or on the imaginary axis. Perhaps most importantly it led to results being found in the frequency domain which are more easily used by the engineer.

## 8.4 Definitions and Loop Transformations

Before presenting some frequency domain results some definitions need to be given and also two useful loop transformations. The results have been derived from either or both Lyapunov's method and functional analysis.

### Definition 5

A system is bounded input – bounded output (BIBO) stable when a bounded input results in a bounded output. Typically a bounded signal is defined in terms of its  $L_2$ -norm, that is

$$\|x(t)\| = \sqrt{\int_0^\infty x^2(t) dt}, \text{ which is finite for a non-periodic BIBO signal.}$$

### Definition 6

The  $A$  matrix of a transfer function  $G(s)$  is Hurwitz if all of its eigenvalues lie in the left hand side of the  $s$ -plane. This will be denoted  $A \in \{A_i\}$

### Definition 7

If the  $A$  matrix is Hurwitz apart from one eigenvalue at  $s = 0$  then  $A \in \{A_0\}$

### Definition 8

If the closed loop system of Figure 3.1 has only eigenvalues in the left hand side  $s$ -plane, when the nonlinearity is replaced by a linear gain  $k$ , this will be denoted by  $A_k \in \{A_i\}$

**Definition 9 – Sector bounded nonlinearity**

A nonlinearity  $n(x)$  which is a real, continuous, single valued, scalar function of  $x$ , with  $n(0) = 0$  belongs to the class  $\{N\}$  if  $k_1 \leq n(x)/x \leq k_2$  for  $x \neq 0$  and is denoted by  $n(x) \in [k_1, k_2]$ .

**Definition 10 – Slope bounded nonlinearity**

A nonlinearity  $n(x)$  which is a real, continuous, single valued, scalar function of  $x$ , with  $n(0) = 0$  belongs to the class  $\{M\}$  if  $m_1 \leq dn(x)/dx \leq m_2$  for  $x \neq 0$  and is denoted by  $n'(x) \in [m_1, m_2]$

**Definition 11 – Monotonic nonlinearity**

A nonlinearity belonging to class  $\{M\}$  which also has  $m_1 \leq [n(x_1) - n(x_2)]/[x_1 - x_2] \leq m_2$  for all finite  $x_1$  and  $x_2 \neq x_1$  belongs to the class  $\{M_m\}$

**8.4.2 Loop Transformations**

Two interesting transformations which are useful for obtaining stability results with the above criteria, and have also been used in the proofs of some of them, are the pole and zero transformations.

Please click the advert



**Nido**

**Luxurious accommodation**

**Central zone 1 & 2 locations**

**Meet hundreds of international students**

**BOOK NOW and get a £100 voucher from voucherexpress**

**Nido Student Living - London**

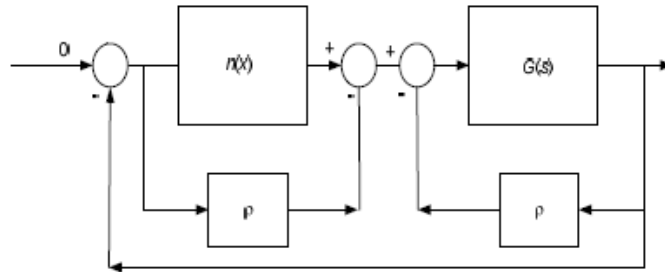
**Visit [www.NidoStudentLiving.com/Bookboon](http://www.NidoStudentLiving.com/Bookboon) for more info.**

**+44 (0)20 3102 1060**

Download free ebooks at [bookboon.com](http://bookboon.com)

### 8.4.2.1 Pole transformation

The closed loop properties for the autonomous system consisting of a nonlinearity  $n(x)$  and plant transfer function  $G(s)$  in the forward path are not changed if a negative feedback path with gain  $\rho$  is placed around the plant and a feedforward path of gain  $-\rho$  is placed around the nonlinearity as shown in Figure 8.1.



**Figure 8.1** Illustration of pole transformation for loop with no input.

The new autonomous loop has linear dynamics  $G_\rho(s) = \frac{G(s)}{1 + \rho G(s)}$  and if  $n(x) \in [k_1, k_2]$  the new nonlinearity  $n_\rho(x) \in [k_1 - \rho, k_2 - \rho]$ . The transformation as well as changing the poles of the linear dynamics changes the sector of the nonlinearity but not the sector width defined by  $k_2 - k_1$  for  $n(x) \in [k_1, k_2]$

### 8.4.2.2 Zero transformation

The added paths of gain  $\sigma$  in the zero transformation are the reverse of those in the pole transformation, that is the feedback path is around the nonlinearity and the feedforward path in parallel with the linear dynamics. The new loop elements are  $G_\sigma(s)$  and  $n_\sigma(x)$  where  $G_\sigma(s) = G(s) - \sigma$  and  $n_\sigma(x) \in \left[ \frac{k_1}{1 + \sigma k_1}, \frac{k_2}{1 + \sigma k_2} \right]$

## 8.5 Frequency Domain Criteria

Several frequency domain criteria are given in the following sections.

### 8.5.1 Popov Theorem

A sufficient condition for the autonomous system of Figure 3.1 to be asymptotically stable with  $A_k \in \{A_1\}$  for  $k = 0$  and  $n(x) \in [0, k]$  is that there exists a non-negative  $q$  such that

$$\operatorname{Re}\{(1 + j\omega q)G(j\omega)\} + k^{-1} \geq \delta > 0 \quad (8.10)$$

where  $\delta$  is an arbitrary small constant. When  $A_k \in \{A_0\}$  for  $k = 0$  then the result is true for  $n(x) \in [\varepsilon, k]$  where  $\varepsilon$  is a small positive constant. Equation (8.10) can be written as

$$\operatorname{Re}(G(j\omega)) > -k^{-1} + q\omega \operatorname{Im} G(j\omega) \quad (8.11)$$

If the real and imaginary axes of the Nyquist plot are denoted by  $x$  and  $y$  then this means that for stability the Nyquist plot of  $G(j\omega)$  should lie to the right of the straight line

$$x = -k^{-1} + q\omega y$$

which is a function of the frequency  $\omega$ , so the line has to be changed for each frequency on  $G(j\omega)$ . The result is usually interpreted graphically in the frequency domain using the Popov locus

$$\tilde{G}(j\omega) = \operatorname{Re} G(j\omega) + j\omega + j\omega \operatorname{Im} G(j\omega)$$

The theorem can then be stated as, 'A sufficient condition for the system to be asymptotically stable is that the Popov locus  $\tilde{G}(j\omega)$  lies to the right of a straight line with finite slope passing through the point  $(-k^{-1}, 0)$ , provided  $\lim_{s \rightarrow \infty} G(s) > -k^{-1}$

The slope of the line is  $q^{-1}$ .

The graphical interpretation of the Popov criterion is shown in Figure 8.2 for two different  $\tilde{G}(j\omega)$  loci labelled G1 and G2. Assuming the transfer functions are such that the linear system will be stable for gains,  $k$ , in the sector  $k \in [0, k_H]$ , which is known as the Hurwitz sector, then it is seen that the Hurwitz and Popov sectors are the same for G2 but the latter is smaller for G1.

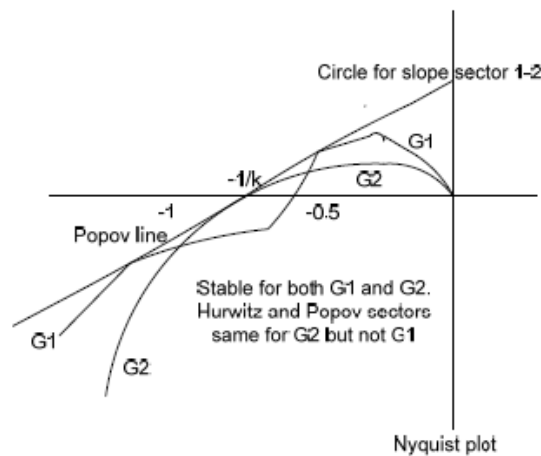


Figure 8.2 Illustration of the Popov criterion.

### 8.5.2 The Circle Criterion

There are two forms of a circle criterion which can be derived for a sector bounded nonlinearity from the Popov criterion. One is the case when  $q$  is taken equal to zero and the criterion is then usually referred to as 'the circle criterion'. For this case it has further been shown to be true for BIBO stability. It is shown in reference [8.1] that if a pole transformation is applied to the autonomous system, which results in a sector transformation for the nonlinearity and this is taken as  $n(x) \in \{N\}$  then the Popov criterion for a stable  $G(s)$  is satisfied if the Nyquist plot of  $G(j\omega)$  does not enter the circle with centre at the point  $-\frac{1}{2}\left(\frac{1}{k_1} + \frac{1}{k_2}\right)$ ,  $\frac{1}{2}q\omega\left(\frac{1}{k_1} - \frac{1}{k_2}\right)$  and which passes through the points  $(-k_1^{-1}, 0)$  and  $(-k_2^{-1}, 0)$ . The dependence of the circle on  $\omega$  is sometimes not a problem since  $q$  can be chosen large enough to take the circle centre well into the third quadrant for any relevant values of  $\omega$ . If  $q$  is taken equal to zero, however, which is a special and more restrictive case of the Popov criterion, then the result is the following theorem.



**Theorem 1 – The Circle Criterion**

The system will be asymptotically stable for  $n(x) \in \{N\}$  and a linear transfer function  $G(s)$ , where the circle  $D$  has its centre on the negative real axis of the Nyquist diagram and passes through the points  $(-k_1^{-1}, 0)$  and  $(-k_2^{-1}, 0)$  if

(i) for  $0 < k_1 < k_2$  the Nyquist plot of  $G(j\omega)$  encircles the circle  $D$ ,  $P$  times counter-clockwise, where  $P$  is the number of poles of  $G(s)$  with positive real parts

(ii) for  $0 = k_1 < k_2$  the Nyquist plot of  $G(j\omega)$  satisfies  $\operatorname{Re} G(j\omega) > -k_2^{-1}$  and  $P = 0$ .

(iii) for  $k_1 < 0 < k_2$  the Nyquist plot of  $G(j\omega)$  lies in the interior of the circle,  $D$ .

The criterion corresponds to the Nyquist criterion for linear systems for (i) with the Nyquist critical point  $(-1, 0)$  becoming the circle  $D$ . Figure 8.3 illustrates the situation for two stable transfer functions  $G_1$  and  $G_2$  whose Nyquist plots close with a semicircle in the right half plane and the circle is for a nonlinearity in the sector  $[1, 3]$ . The closed loop nonlinear system with  $G_1$  is stable according to the criterion but no conclusion can be reached about  $G_2$ . It should be noted that as  $k_1$  tends to zero the circle tends to a vertical straight line through  $-k_2^{-1}$ . Obviously, a disadvantage of the criterion for autonomous systems, since it is a special case of the Popov criterion, is that it produces more conservative results.

Please click the advert

I joined MITAS because  
I wanted **real responsibility**

The Graduate Programme  
for Engineers and Geoscientists  
**Maersk.com/Mitas**



**Month 16**

I was a construction  
supervisor in  
the North Sea  
advising and  
helping foremen  
solve problems

Real work  
International opportunities  
Three work placements



 **MAERSK**

Download free ebooks at [bookboon.com](http://bookboon.com)



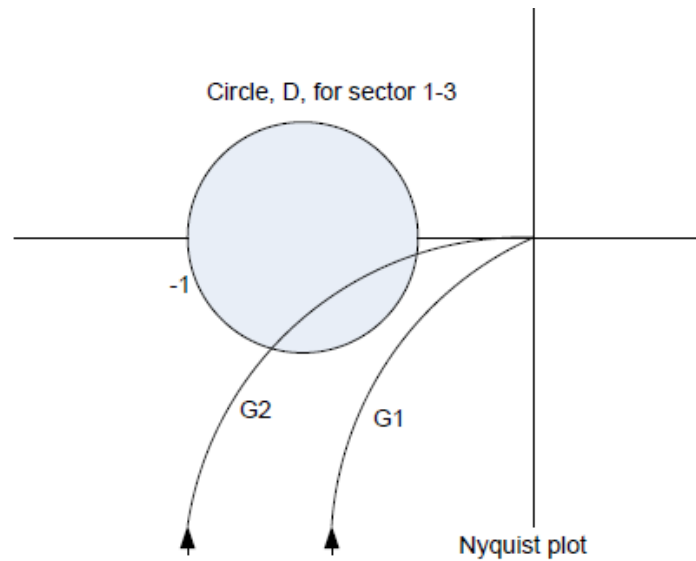


Figure 8.3 Illustration of the circle criterion

It is interesting to compare this result with the DF method as for any nonlinearity  $n(x) \in \{N\}$  it was shown in section 3.3 that  $N(a) \in \{N\}$  so the DF for the class  $N$  of single valued nonlinearities is the diameter of the circle of Fig 8.3. Thus for 'DF stability'  $G(j\omega)$  should not cut the diameter of the circle. Aizermann conjectured that 'DF stability' was correct but the 'so-called' counter example of Fitts given in section 5. 5 shows this to be untrue. The feature of the Fitts example which breaks 'DF stability' down is that  $G(j\omega)$  exists in both the second and third quadrants (defined in a counter clockwise direction from the first quadrant within the positive axes). It is interesting to note that if  $G(j\omega)$  is confined only to the third and fourth quadrants then 'DF stability' can be proved by the more general circle criterion discussed at the start of this section. Since this is a typical region for many Nyquist loci, that is the plant transfer function has a phase lag between 0 and  $180^\circ$ , this result may be regarded as the small phase shift stability theorem.

### 8.5.3 The Off-axis Circle Criterion

The off-axis circle criterion applies when the nonlinearity is monotonic. Its applicability is therefore restricted but it can give less conservative results than the other criteria.

#### Theorem 2

The autonomous system with  $G(s)$  such that  $A_{m1}$  and  $A_{m2} \in A_1$  and  $n(x) \in \{M_m\}$  with  $n'(x) \in [m_1, m_2]$  is asymptotically stable if the Nyquist locus  $G(j\omega)$  for  $\omega \in [0, \infty]$  does not enter the circle  $D'$ , where  $D'$  is a circle which passes through the points  $(-m_1^{-1}, 0)$  and  $(-m_2^{-1}, 0)$  with centre anywhere in the Nyquist plane.

#### Remark

For the case of  $m_1=0$  the circle becomes a straight line through  $(-m_2^{-1}, 0)$ , so that for stability the Nyquist locus must lie to the right of this line. Figure 8.4 illustrates the situation showing the loop stable for G1 but not for G2. However stability for G2 can be proved by choosing a larger radius for the circle and moving its centre farther away from the x axis.

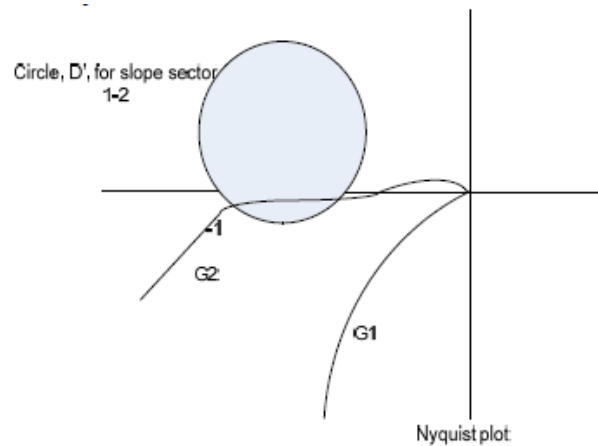


Figure 8.4 Illustration of the off-axis circle criterion.

### Remark

Since the diameter of the circle can be made very large the limit of the chord is the line between the points  $(-m_1^{-1}, 0)$  and  $(-m_2^{-1}, 0)$ . Thus for a monotonic nonlinearity instability can only occur if the Nyquist locus exists in both the second and third quadrants. This again shows why a  $G(j\omega)$  like that in the Fitts counterexample breaks down a DF stability type result based on the slope bounds of the nonlinearity which was conjectured by Kalman.

## 8.6 Examples

Several examples are presented in this section to allow comparisons of the various approaches.

### Example 1

Consider a system with an open loop transfer function of  $G(s) = k(1-s)/2(1+s)$ . Substituting  $s = j\omega$ , with  $k = 1$ , gives

$$G(j\omega) = \frac{1-j\omega}{2(1+j\omega)} = \frac{(1-j\omega)^2}{2(1+\omega^2)} = \frac{(1-\omega^2)}{2(1+\omega^2)} - \frac{j\omega}{(1+\omega^2)}$$

so that the Popov locus is

$$\tilde{G}(j\omega) = \frac{1-\omega^2}{2(1+\omega^2)} - \frac{j\omega^2}{1+\omega^2} \text{ which on writing } \tilde{G}(j\omega) = \tilde{X} + j\tilde{Y} \text{ and eliminating } \omega \text{ from the}$$

expressions for  $\tilde{X}$  and  $\tilde{Y}$  gives the straight line  $\tilde{Y} = \tilde{X} - 0.5$ . The locus starts at  $(0.5, 0)$  for  $\omega=0$  and finishes at  $(-0.5, -1)$  for  $\omega=\infty$ . Figure 8.5 shows the Nyquist and Popov loci, with the latter drawn for positive frequencies only. It is clear that a line can be drawn through the origin with slope approximately 2.0 so that the Popov locus lies to its right. This means a  $k$  equal to infinity, however the criterion also requires  $\lim_{s \rightarrow \infty} G(s) > -k^{-1}$ , which gives  $-2^{-1} > -k^{-1}$ , that is  $k < 2$ . The Hurwitz sector is  $(-2, 2)$  so that the Hurwitz and Popov criteria are the same for positive gains.

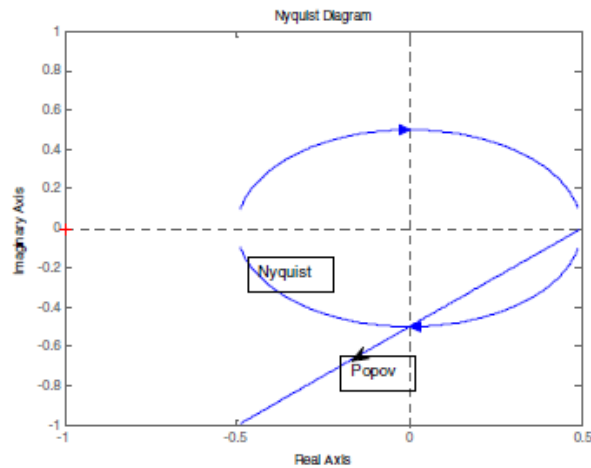


Figure 8.5 Nyquist and Popov Plots

## Example 2

Consider a nonlinear feedback system with  $G(s) = -1/(1-sT_1)(1+sT_2)$  and a nonlinear element in the sector  $n(x) \in (k_1, k_2)$ . It is required to find conditions for stability of the closed loop system. The characteristic equation for the linear system with a loop gain of  $k$  is  $sT_1T_2 + s(T_1 - T_2) + k - 1 = 0$  so that the Hurwitz sector is  $(1, \infty)$  provided  $T_1 > T_2$ . For the nonlinear system the Popov criterion cannot be applied directly as  $G(s)$  is not stable, that is  $A_k \notin \{A_0\}$  for  $k=0$ . However if a pole transformation is used with a gain  $\rho$  then  $G_\rho(s) = 1/(s^2T_1T_2 + s(T_1 - T_2) + \rho - 1)$  and  $n_\rho(x) \in [k_1 - \rho, k_2 - \rho]$ .

Now  $G_\rho(s)$  is stable provided  $T_1 > T_2$  and  $\rho > 1$  and since the Popov locus of  $G_\rho(j\omega)$  is entirely within the lower half plane the Popov sector for  $n_\rho(x)$  is  $(0, \infty)$  which gives stability for  $n(x) \in (1, \infty)$ , provided  $T_1 > T_2$ , in agreement with the Hurwitz sector. For  $n(x) \in (1, \infty)$  the circle D for the circle criterion is a unit circle with centre at  $(-0.5, 0)$ . It is easy to show that the Nyquist plot of  $G(s)$  will lie within this circle for all  $T_1, T_2$  so that stability cannot be proved.

## Example 3

Consider a system with  $G(s) = \frac{s+0.8}{s(s+0.1)(s^2+0.3s+9)}$  with  $n(x) \in \{N\}$  and  $\{M\}$  such that  $n(x) \in \{0, k\}$  and  $n'(x) \in \{0, k\}$ . The Nyquist and Popov loci for the transfer function are shown in Fig 8.6 from which it can be seen that the Hurwitz sector is greater than the Popov sector. As  $\omega \rightarrow 0$ , the real part of  $G(j\omega)$  tends to  $-7.81$  so that the locus will lie to the right of a vertical straight line passing through  $(-7.81, 0)$ , giving a gain for stability of  $k < 1/7.81 = 0.128$ . A Popov line drawn to keep the Popov locus to the right intersects the negative real axis at around  $-3.40$  so to satisfy the Popov criterion  $k < 0.29$ . Similarly a line drawn to keep the Nyquist locus to the right will intersect the negative real axis at  $-0.48$  giving  $k < 2.08$  for stability. Finally for the linear system the maximum gain for stability from the Hurwitz sector is  $2.61$ .

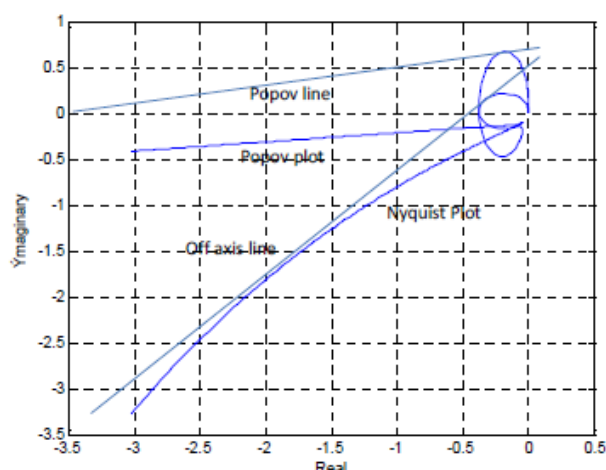


Figure 8.6 Frequency Responses for Example 3

## 8.7 Conclusions

In this chapter some absolute stability results have been presented primarily those using frequency domain criteria. Their major advantages are that they allow comparisons with the approximate DF method and because of their formulation allow stability to be proven for classes of nonlinearity, rather than a specific one, so that they are robust to changes in the nonlinearity. Their disadvantages are that results of their application can prove conservative for a specific nonlinearity and they are restricted in their applicability to a feedback loop with a single nonlinear element.

## 8.8 References

Please click the advert

# Brain power

By 2020, wind could provide one-tenth of our planet's electricity needs. Already today, SKF's innovative know-how is crucial to running a large proportion of the world's wind turbines.

Up to 25 % of the generating costs relate to maintenance. These can be reduced dramatically thanks to our systems for on-line condition monitoring and automatic lubrication. We help make it more economical to create cleaner, cheaper energy out of thin air.

By sharing our experience, expertise, and creativity, industries can boost performance beyond expectations. Therefore we need the best employees who can meet this challenge!

**The Power of Knowledge Engineering**

Plug into The Power of Knowledge Engineering.  
Visit us at [www.skf.com/knowledge](http://www.skf.com/knowledge)

# 9 Design of Nonlinear Control Systems

## 9.1 Introduction

In agreement with the situation for books on linear control systems the early chapters of this book have necessarily been devoted to analysis methods rather than design. In practice the design process involves consideration of the approach to be used to meet the given specifications, the selection of the components to be used and their modelling and finally the use of analytical methods and probably simulation to show that the design meets the specifications. This is normally an iterative process requiring a return to earlier stages in the process, making changes and then continuing. The difficulty if any components are nonlinear is, as has been seen, that the behaviour is dependent on the magnitude and type of input signals to which the system is subjected, no necessary and sufficient conditions for stability assessment are known and no general system analysis method is available.

Thus if the system specification requires the input step response to have certain properties, then it is not sufficient to know that this will be the case for a step input of a single magnitude, as with a linear system, but one must determine that the requirements will be satisfied for all step inputs within the allowable range. Also if the system is subject to multiple inputs it is not sufficient to determine the behaviour for the inputs separately since nonlinear interactions between them may cause unexpected behaviour. The behaviour of high order nonlinear systems is dependent on the region of the state space in which operation takes place, as has been demonstrated for a second order system by use of the phase plane. Once a system has been designed simulation provides an excellent tool for checking the performance within the allowable state space region of operation, however, if it is done without theoretical understanding it can prove inefficient and possibly erroneous if the simulations fail to cover appropriate scenarios.

One might argue that the first objective of a design should be to try and add compensation to make the system linear, so as to allow easy analysis, but this may not be possible or even desirable. The advantage of linearity is that the response is independent of the input amplitude but it may only be possible to have linear behaviour between a specific input and output. A very common form of nonlinearity is plant actuator saturation, for example motor torque saturation in a speed or position control system. It might be possible to avoid this by using a larger motor but if the system satisfies the specifications with a smaller motor then the consequence of changing to a larger one is a more costly design. An objective of any engineering design is to meet the specifications at the lowest cost. The specifications are paramount and as well as performance and reliability issues they may cover other aspects such as operating costs, weight or volume, etc. and environmental concerns. In some cases when the actuator to a plant saturates it can cause problems as the feedback loop is then essentially operating open loop. One extreme case is when the plant model is an unstable transfer function.

In many cases where nonlinearity occurs it may be possible to do a control design based on an approximate linear or linearized system model and then check by simulation that the effects of the nonlinearity still allow the required performance to be met. Such a design may involve parameter selection for fixed controllers like PID, or one of several state space design methods. The latter may use state variables estimated using observers and although unlike for linear systems there is no rigorous justification provided by a separation principle the procedure is effective. Today most controllers are implemented using microprocessors so that it is just as easy to implement a nonlinear control law as a linear one. When should this be done and why? What are the advantages and disadvantages?

In the following sections some design methods for nonlinear systems are mentioned and hopefully these will go some way to answering these questions. The topics, however, are only covered briefly to give a 'flavour' of what is available and also because to do otherwise would require the introduction of additional theoretical concepts. Complete texts are available on most of the topics and references will be given in the appropriate sections. There are also several books that give a good overview of many of the topics considered in this chapter, such as references 9.1 and 9.2. The latter, reference 9.2, because of the author's industrial background is particularly interesting in offering a much more practical perspective than is often the case.

## 9.2 Linearization

Nonlinear systems are often designed using linear methods based on an approximate linear model. Simulations may show that satisfactory performance may be obtained from a design using one linear model but in some situations this will not be possible due to the large changes in the best linear model with respect to the system operating point. Linear controllers may then be designed for the different operating point models, or linearizations. The controller parameters are then 'scheduled' in some manner based on knowledge of the operating point. Many approaches are possible which can be confirmed to work satisfactorily in simulations although there is sometimes a lack of rigorous theory to support the techniques.

The approach is particularly common in the aerospace industry where a common approach is to employ state feedback with the gains scheduled according to the specific operating point. The design is typically done off line in that a linear mathematical model is obtained for each operating point and then the required controller gains calculated and stored ready for implementation around that specific operating point. Various methods are also used to avoid discontinuous changes of controller gains with change in the operating point. Two good introductory references for work in this field are [9.3, 9.4]. If a new mathematical model is identified during operation then this provides a standard form of adaptive control. All control designs typical involve considerations of system performance under plant changes, and how severe these may be expected to be will affect whether a robust design can be accomplished by a single controller or whether scheduled controllers or an adaptive control is required.

Many processes which need to be controlled in the chemical industry often have quite nonlinear dynamics so these are sometimes controlled by PID controllers which allow the three PID terms to be scheduled according to the operating point. In recent years there has also been a significant amount of work on the topic of multiple models for use in both modelling and control [9.5]




### 9.3 Frequency Response Shaping

In classical controller design methods for linear systems it is a common procedure to add known forms of compensator, such as phase lead or lag, to achieve a desired frequency domain property, such as a minimum allowable gain or phase margin. In such designs the requirement is then to obtain suitable controller parameters so the system meets the specifications. This is again a satisfactory approach for a nonlinear system and in terms of the DF approach, assuming a single nonlinear element, becomes one of shaping the open loop frequency response locus so that it lies at an appropriate distance from the  $C(a)$  locus. The  $C(a)$  locus may be regarded as a 'moving  $(-1, 0)$ ' point on the Nyquist plot so a gain and phase margin may be defined for each value of  $a$ .

Typically a design will involve shaping the frequency response locus to ensure that the value of the gain and/or phase margin for all  $a$  is never less than a given value.

Apart from the obvious way of doing this by comparing the location of the  $C(a)$  and  $G(j\omega)$  loci on a Nyquist plot, or on a Bode plot, particularly for a single valued nonlinearity, another approach is to make use of methods for determining limit cycles. Using such an approach the amount of gain or phase shift, the latter by adding a time delay, which may be added to the loop before a limit cycle results can be calculated.

Please click the advert




## Are you considering a European business degree?






**LEARN BUSINESS** at university level. We mix cases with cutting edge research working individually or in teams and everyone speaks English. Bring back valuable knowledge and experience to boost your career.

**MEET** a culture of new foods, music and traditions and a new way of studying business in a safe, clean environment – in the middle of Copenhagen, Denmark.


**ENGAGE** in extra-curricular activities such as case competitions, sports, etc. – make new friends among CBS' 18,000 students from more than 80 countries.



**Copenhagen Business School**  
HANDELSHØJSKOLEN

See what we look like and how we work on [cbs.dk](http://cbs.dk)



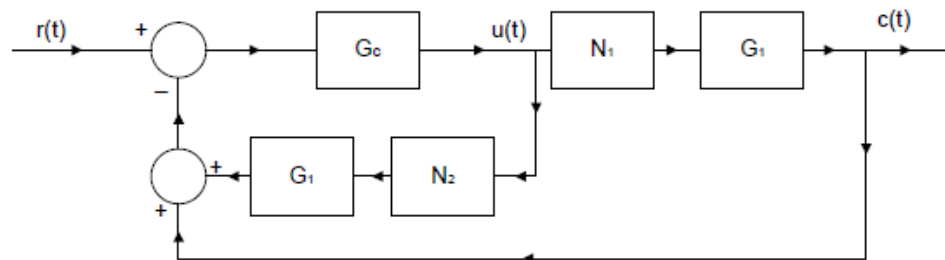
Download free ebooks at [bookboon.com](http://bookboon.com)



## 9.4 Nonlinear Compensation

Given the block diagram for a feedback system with a nonlinear element, possibilities may exist, dependent upon the location of the nonlinearity, for compensation using another nonlinearity. For example, if nonlinearity exists at the input to the plant then it may be possible to include a nonlinearity in the controller output to make the loop behaviour 'more linear'. Typically this might be used to provide a higher gain for small signals to compensate for a valve dead zone. From purely block diagram considerations it is possible to include another path to the system output in parallel with the plant, but in practice this is not physically possible. However, the closed loop design problem may be made linear by taking such a path and connecting it to the feedback signal as shown in Figure 9.1. If in the figure the nonlinear element  $N_2$  is taken equal to  $1-N_1$  then it is easy to show that the closed loop characteristic equation is linear as is the transfer function from the input,  $R$ , to the control signal,  $U$ , which is

$$\frac{U(s)}{R(s)} = \frac{G_c}{1 + GG_c} \quad (9.1)$$



**Figure 9.1** A Nonlinear Compensation Method

Sometimes nonlinearity occurs in a minor feedback path, for example, when the damping is a nonlinear function of speed, unlike viscous friction which is assumed to vary linearly with speed. Obviously, if speed is measured then feeding back a nonlinear function of this signal can be used to make the total speed feedback linear. Practically, however, this means knowing the form of the nonlinear speed damping and may also require an additional sensor for speed measurement or the implementation of a speed estimator.

A very common use of nonlinear compensation is for the prevention of integral windup, which is provided with most commercial PID controllers. As mentioned previously saturation always eventually occurs in a plant, whether it be the maximum torque from a motor, the maximum power from a heater, the maximum flow through a valve and at this stage further increase in the controller output will not increase the level of the system output, allowing of course for any possible delay in the response. If the controller contains an integral term then its output will increase in magnitude and not start to decrease until its input changes sign. This can cause large overshoots in the system response and in the case of an unstable plant, since the loop is basically open, possibly catastrophic behaviour. Thus, integral windup protection is achieved by stopping the integration when the controller output signal reaches a certain level. Today, since PID controllers use microprocessors various software algorithms are used to implement the procedure, but in the days of analogue controllers a simple approach was to incorporate a nonlinear feedback path around the integrator. Another approach to avoiding integral windup [9.6] is simply to switch on the integrator only when near to the steady state, which in some ways seems more logical as the integration normally has a destabilising effect on the loop and the only reason it is usually included is to eliminate steady state errors. A slight difficulty is in obtaining a suitable algorithm to implement the switching on

and off conditions without causing transient problems.

## 9.5 Compensation using DF Models

In the previous section it was assumed that any nonlinearity in the control system could be represented in a simple block diagram structure. In practice, such modelling may not be easily achieved or may not be of relevance for the approach because of the complexity of the plant model structure with multiple dynamic and nonlinear elements. In this case a possible approach is to make use of a frequency dependent DF model,  $N(a, \omega)$ , for the plant, which may be obtained from either a mathematical model of the plant or possibly frequency response identification testing. As long as the frequency response shows a dominant output at the input frequency the model should be satisfactory for use in compensator design studies. Two approaches are possible, the first simply being a DF implementation of the structure of Figure 9.1, that is if the nonlinear plant has the model,  $N_1(a, \omega)$ , then the DF model fed back around the controller  $N_2(a, \omega)$ , should be such that  $N_1(a, \omega) + N_2(a, \omega) \approx G(j\omega)$  a linear model. To DF accuracy the open loop is then essentially linear, so that linear compensation can be used.

An alternative approach is to place a series compensator before the plant which has a DF model  $N_2(a, \omega)$  such that the series combination  $N_2(a_2, \omega)$  and  $N_1(a_1, \omega)$  are approximately equal to  $G(j\omega)$  a linear model. The different subscripts on the  $a$ 's have been used to denote that the amplitudes into the DF models will not be the same. More details on the actual approach to achieving this design can be found in reference 9.7. The approach has also been extended for use with relay autotuning (see chapter 7) to implement nonlinear terms in a PID controller. When the relay autotuning approach is used for a nonlinear plant then the estimated critical point will change as the relay output amplitude is changed. This not only enables one to estimate the 'degree of nonlinearity' in the plant but also enables the design of a nonlinear PID to produce roughly a linear input output response [9.8].

## 9.6 Time Optimal Control

Early work in time optimal control was done for second order systems by making use of the phase plane approach. However, the topic developed rapidly in the 50's and a large number of papers on the subject were published in the 60's. Although much of the early work was done in the USSR, major developments started in the USA in the 60's primarily driven by the space program where an important requirement was to develop control strategies using minimal fuel, a problem which is closely related to time optimal control. The concepts used are based on classical Hamiltonian mechanics and generally are known as Pontriagin's maximum principle or the Bellman-Hamilton-Jacobi approach. The books by McCausland [9.9], Jacobs [9.10], and Glad and Ljung [9.1] give good introductions and more advanced books are those by Ryan [9.11], Athans and Falb [9.12] and Lewis [9.13]. The mathematical theory of these approaches is not easy so here a simple introduction to the topic is provided by way of the phase plane for a double integrator plant. The solution to the problem using the classical Hamiltonian approach is given in chapter 6 of reference 9.9.

When considering the initial condition response of a second order system in section 2.3.2 it was seen that the phase plane plot for a closed loop system consisting of a relay and a double integrator plant transfer function consisted of parabolas. This corresponds physically to a position control system, assuming a constant inertial load and no friction, being driven at constant acceleration,  $K$ , with the equation of the parabolas being


$$x_2^2 - x_{20}^2 = \pm 2K(x_1 - x_{10}) \quad (9.2)$$

where  $x_1$  is position and  $x_2$  velocity. If the input to the relay is  $-x_1 - \lambda x_2$  the switching between parabolas corresponding to positive and negative acceleration (deceleration) takes place on the line  $-x_1 - \lambda x_2 = 0$  in the phase plane.

It is easily appreciated that the fastest possible way to the origin from an initial condition of  $(-h, 0)$  is to have the maximum acceleration of  $+K$  to a distance  $-h/2$  from the origin followed by the maximum deceleration of  $-K$  to the origin. The final trajectory to the origin is the parabola  $x_2^2 = -2Kx_1$ , or if the motion starts from  $(h, 0)$ , rather than  $(-h, 0)$ , the final trajectory is  $x_2^2 = 2Kx_1$ . Thus to achieve a minimum time response in the ideal case these two parts of parabolas should form the switching curve not the line  $-x_1 - \lambda x_2 = 0$ . A minimum time response will therefore be obtained if the input to the relay is  $-x_2|x_2| - 2Kx_1$ . This has been called a  $V|V|$  system [9.14]. It is, however, often implemented by having the switching boundary as  $-x_2 + \sqrt{2K} \operatorname{sgn}(-x_1)\sqrt{|-x_1|} = 0$ , when it is known as the SERME (sign error root modulus error) system.

The optimal solution is seen to be dependent on knowing the system model, in this case the parameter  $K$  assuming negligible friction. If  $K$  is estimated, however, the switching boundary can be adjusted by some adaptive strategy.

Please click the advert



**The financial industry needs a strong software platform**  
That's why we need you

SimCorp is a leading provider of software solutions for the financial industry. We work together to reach a common goal: to help our clients succeed by providing a strong, scalable IT platform that enables growth, while mitigating risk and reducing cost. At SimCorp, we value commitment and enable you to make the most of your ambitions and potential.

Are you among the best qualified in finance, economics, IT or mathematics?

**Find your next challenge at**  
**[www.simcorp.com/careers](http://www.simcorp.com/careers)**



[www.simcorp.com](http://www.simcorp.com)

MITIGATE RISK | REDUCE COST | ENABLE GROWTH

## 9.7 Sliding Mode Control

The phenomenon of sliding was encountered in Example 2 of Chapter 2, where a phase plane plot was constructed for a second order system containing a relay with dead zone and hysteresis. What happens, as explained in the example, is that the dynamics of the motion changes when the relay switches and it is possible that at both sides of the switching line the resultant dynamic motion is towards the switching line, producing a motion which is directed with rapid switching along the switching line. The interesting feature recognised about this sliding motion is that once the sliding motion is started the trajectory is defined and may not change even with some changes in the system dynamics. Sliding mode control is a particular form of variable structure control, where the feedback law is selected from several possibilities by decision rules. Variable structure control systems (VSCS) were first studied in the USSR [9.15]. A second order system will again be considered in the next section before a few comments on the more general situation.

### 9.7.1 A Second Order System

Consider again the same system as Example 2, Chapter 2, but in this case with the relay nonlinearity replaced by an ideal relay. Again the feedback to the relay is assumed to be  $-x_1 - \lambda x_2$  and the transfer function of the plant after the relay to be  $K/s^2$ . The Simulink diagram is shown in Figure 9.2 with  $\lambda = K = 1$ . Assuming an initial condition of  $(-x_{10}, 0)$  then since the relay output  $u = \text{sgn}(-x_1 - \lambda x_2)$  the initial value of  $u$  is positive and the equation of motion is  $\ddot{x}_2 = 2K(x_1 - x_{10})$  which meets the switching line  $x_1 + \lambda x_2 = 0$  at say  $(x_{1s}, x_{2s})$ . The motion after switching is then described by  $\ddot{x}_2 - x_{2s} = -2K(x_1 - x_{1s})$  and this will move back to the switching line if its slope is less than  $-1/\lambda$ . That is if  $-K/x_2 < -1/\lambda$ , which gives

$$x_2 < \lambda K. \quad (9.3)$$

This condition for ensuring sliding is known as the reachability condition.

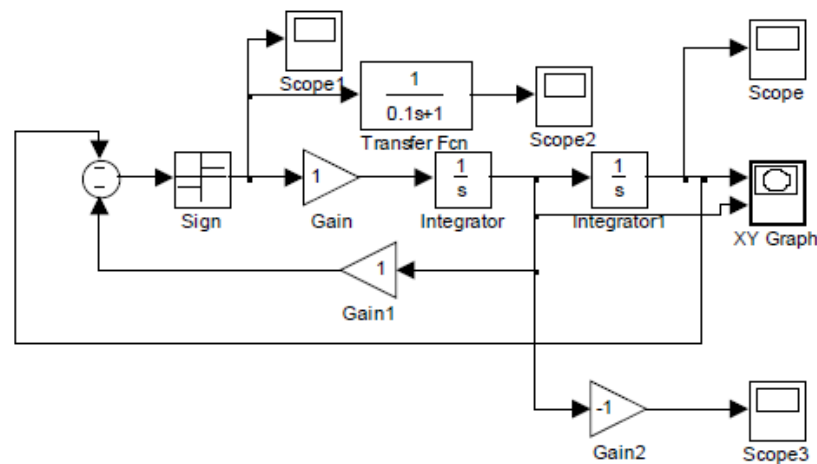


Figure 9.2 Simulink diagram for a system with a sliding mode.

It is normally given in a more general form since if the input to the relay is taken as  $s$  so that  $s = -x_1 - \lambda x_2 = 0$  is the switching line, the control signal,  $u$ , is given by

$$u = \text{sgn}(s) \quad (9.4)$$

then if the switching line is to be reached it requires  $\lim_{s \rightarrow 0^+} \dot{s} < 0$  and  $\lim_{s \rightarrow 0^-} \dot{s} > 0$ , which can be expressed as the single reachability condition

$$s\dot{s} < 0 \quad (9.5)$$

As the motion in the phase plane is symmetrical the actual reachability condition using equation (9.3), or as can easily be shown from equation (9.5), is

$$|x_2| < \lambda K. \quad (9.6)$$

When ideal sliding takes place the resulting motion is along the switching line and is governed by the differential equation  $s = 0$ , which in the case of the example is  $\frac{dx_1}{dt} + \frac{1}{\lambda}x_1 = 0$ . The control action required to bring about this motion is discontinuous and not defined on the switching line. However, in addition when sliding takes place  $\dot{s} = 0$  and therefore  $s = -\dot{x}_1 - \lambda Ku = 0$ , giving

$$u = -\dot{x}_1 / K\lambda = -x_2 / \lambda K \quad (9.7)$$

which is known as the equivalent control.

Figure 9.3 shows the phase plane response for the system of Figure 9.2 from the initial condition  $(-1, 0)$ . The switching line  $x_1 + x_2 = 0$  is first met with  $x_2 < 1$ , the reachability condition of equation (9.3) for  $\lambda = K = 1$ , so sliding takes place. Figure 9.4 shows the system output,  $x_1$ , (Scope 1);  $-x_2$  (Scope3) and the approximate equivalent control (Scope 2) obtained by filtering,  $u$ , which is shown in Figure 9.5 continuously switching during sliding. The approximate equivalent control can be seen to become equal to  $-\dot{x}_1 = -x_2$  during sliding as given by equation (9.7) for  $\lambda = K = 1$ .

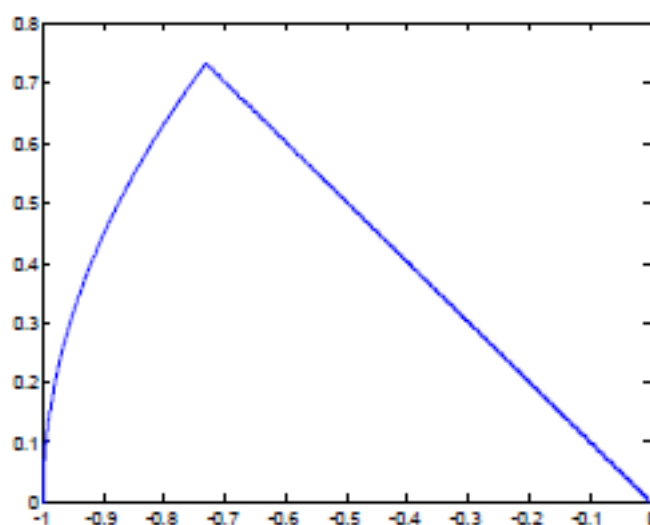


Figure 9.3 Motion in the phase plane for input from  $(-1, 0)$

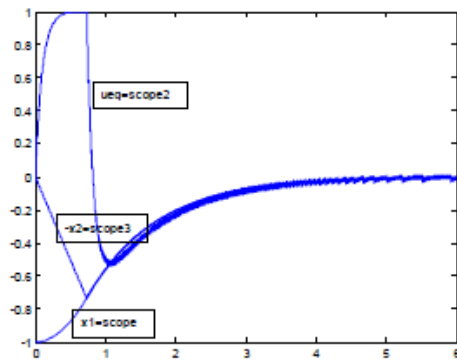
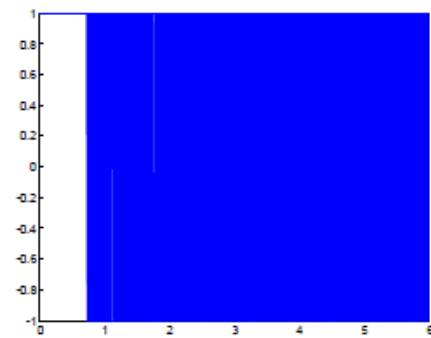


Figure 9.4 Signals in the Simulation

Figure 9.5 Control signal  $u$ 

Let it now be assumed that the plant model is not an ideal double integrator but has some nonlinear damping. This is illustrated in Figure 9.6 where the nonlinear damping has been made quite significant, whereas in practice one would not expect such a significant change from the assumed model, so that the change in the response on the phase plane is clearly seen in Figure 9.7. The switching line is reached on a different trajectory but once it is attained sliding takes place as before.

Please click the advert

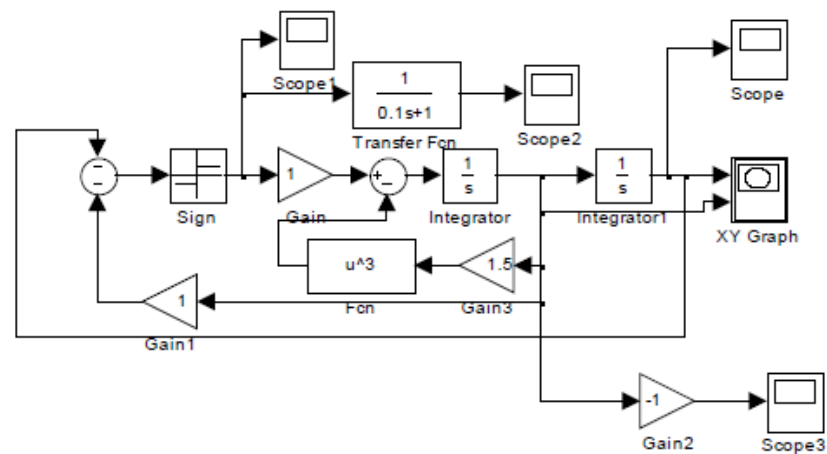
# What do you want to do?

No matter what you want out of your future career, an employer with a broad range of operations in a load of countries will always be the ticket. Working within the Volvo Group means more than 100,000 friends and colleagues in more than 185 countries all over the world. We offer graduates great career opportunities – check out the Career section at our web site [www.volvogroup.com](http://www.volvogroup.com). We look forward to getting to know you!

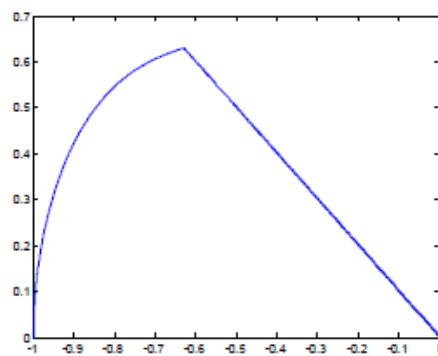
**VOLVO**  
AB Volvo (publ)  
[www.volvogroup.com](http://www.volvogroup.com)

VOLVO TRUCKS | RENAULT TRUCKS | MACK TRUCKS | VOLVO BUSES | VOLVO CONSTRUCTION EQUIPMENT | VOLVO PENTA | VOLVO AERO | VOLVO IT  
VOLVO FINANCIAL SERVICES | VOLVO 3P | VOLVO POWERTRAIN | VOLVO PARTS | VOLVO TECHNOLOGY | VOLVO LOGISTICS | BUSINESS AREA ASIA

Download free ebooks at [bookboon.com](http://bookboon.com)



**Figure 9.6** Simulink Diagram for System with Nonlinear Damping.



**Figure 9.7** Phase plane response for system of Figure 9.6.

The reachability condition for this example, from equation (9.6), is  $|x_2| < 1$  and it may be that reachability is required for a larger region in the phase space. Consider the control signal  $u$  to be a summation of state feedback and the signum function, so that

$$u = k_1 x_1 + k_2 x_2 + \rho \operatorname{sgn}(s)$$

$$\text{Then } s\dot{s} = s(-\dot{x}_1 - \lambda\dot{x}_2) = s(-\dot{x}_1 - \lambda K[k_1 x_1 + k_2 x_2 + \operatorname{sgn}(s)])$$

Choosing  $k_1 = 0$  and  $k_2 = -1/\lambda K$  gives  $s\dot{s} = s[-\lambda K\rho \operatorname{sgn}(s)] = -\lambda K\rho |s| = -\eta |s|$  where  $\eta$  is a positive design scalar. Selection of  $s\dot{s} < -\eta |s|$  is known as the  $\eta$ -reachability condition. The control signal

$$u = -(x_2/\lambda K) + \rho \operatorname{sgn}(s) \quad (9.8)$$



which is seen to be the equivalent control of equation (9.7) plus the signum function. In this case the switching surface is reachable from all regions in the phase plane. Figure 9.8 shows the Simulink diagram for this system and Figure 9.9 the response from an initial condition of  $(-6,0)$ . Curve (a) is for the control law of equation (9.4), that is the system of Figure 9.2 and curve (b) for the control law of equation (9.8), that is the system of Figure 9.8. Also shown for the second case is curve (c) the response from  $(-6, 2)$ .

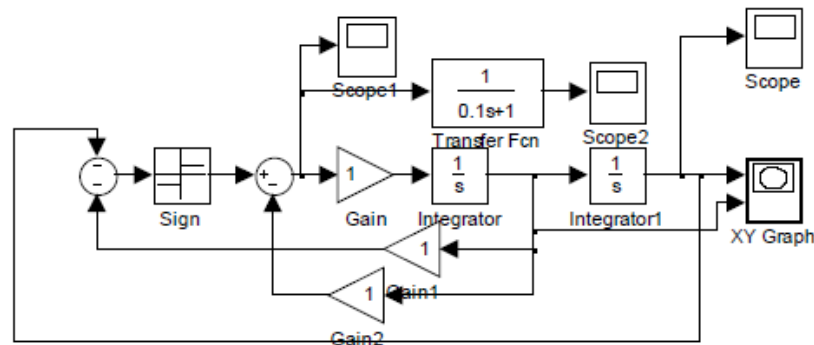


Figure 9.8 System with control law of equation (9.8)

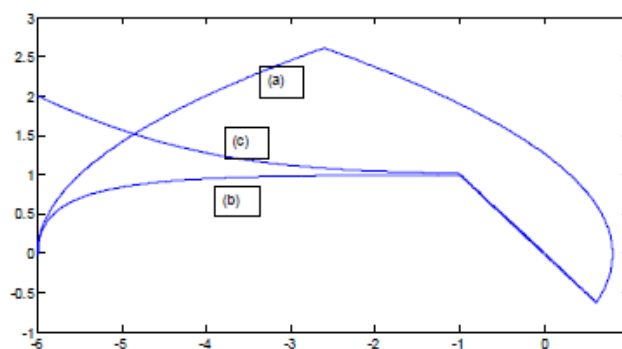


Figure 9.9 Phase plane responses for the two different control laws.

From the considerations of a simple second order system it has been seen that when the reachability condition, given by equation 9.5, is satisfied, sliding is obtained down the switching curve, which of course need not be linear. Further the motion can be regarded as having an equivalent control action to produce a sliding response which is first order, that is reduced by one, from that of the original system.

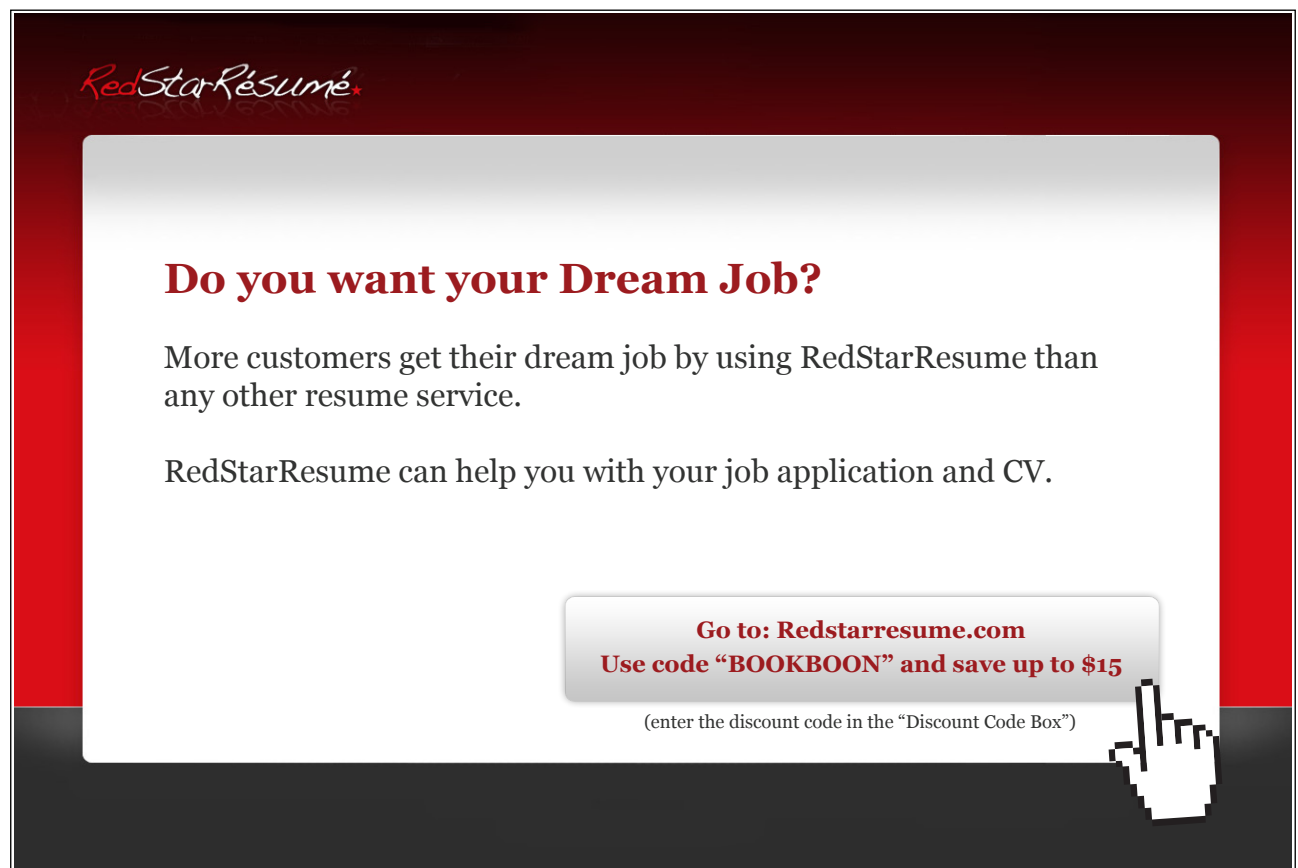
The formulation presented has used the ideal mathematical signum function, which cannot be realised in practice, although it can be nearly ideal in some implementations. However, the effects of very rapid switching may cause practical problems such as rapid wear of components, so the ideal may be approximated by using hysteresis in the switching or using an approximation to the signum function such as  $\text{sgn}(s) \approx \frac{s}{|s| + \delta}$  where  $\delta$  is a small positive constant.

## 9.7.2 Further Comments

The presentation of the previous section has hopefully indicated the objectives of sliding mode control design, namely to design a controller so that the motion will reach, and subsequently remain on a sliding surface, in contrast to a line for a second order system, in the state plane. The dynamics of the system when confined to the surface is described as an ideal sliding motion and for a system with a single input will be reduced in order by one from that of the original system. The controller design for high order systems is not easy to cover in a short presentation of the topic but there are many specialised texts available on the subject such as Utkin [9.16], Zinober [9.17], and Edwards and Spurgeon [9.18].

It is shown in these texts how designs can be achieved using various state space approaches for MIMO as well as SISO systems, what type of plant uncertainty will not affect the sliding motion, and how the ideas can be applied to nonlinear systems and sliding surfaces, not just hyperplanes. Like most state space approaches all the states may have to be available for implementation, so additional instrumentation or observers may be required, although some results are available based only on output feedback. There has also been significant coverage in recent years on the design of observers using sliding mode principles [9.18].

Please click the advert



*RedStarResume.*


### Do you want your Dream Job?

More customers get their dream job by using RedStarResume than any other resume service.

RedStarResume can help you with your job application and CV.

**Go to: [Redstarresume.com](http://Redstarresume.com)**  
**Use code "BOOKBOON" and save up to \$15**

(enter the discount code in the "Discount Code Box")



Download free ebooks at [bookboon.com](http://bookboon.com)

## 9.8 Fuzzy Logic

The concept of fuzzy logic was introduced by Lotfi Zadeh [9.19]. It is perhaps an unfortunate choice of name from a control design viewpoint because a controller designed using fuzzy logic is deterministic having a defined output for a defined input. It is the procedure of producing the output for a specific input which is the 'fuzzy aspect'. This procedure varies slightly according to the approach used but that due to Mamdani [9.20] contains membership functions, fuzzification and defuzzification procedures, and a set of rules. Typically, standard forms for the first three may be available and the designer has only to provide a suitable set of rules. Thus, as stated by Zadeh, it enables a person with no knowledge of control theory to design a controller by writing a few logic type rules, which in principle may be selected based on the observed response of the plant.

The first point to make about the procedure is that in general the controller designed will be nonlinear, so the wisdom of doing this if the plant is essentially linear is debatable. Secondly, it seems to be a fairly common practice to write the logic rules for a controller in the error channel in terms of the error and its derivative. Thus the controller becomes a nonlinear controller of the form  $f(e, \dot{e})$ . It is not easy to see what advantage a controller of this PD form may have compared with the less general PD form  $f_1(e) + f_2(\dot{e})$ , a classical PD controller with nonlinear P and D terms. Many papers describing a fuzzy logic controller end up with a nonlinear one yet fail to check the response variation with the amplitude of the input signal.

A very good discussion on fuzzy logic controllers can be found in the results of the DynLAB project of reference 9.21. In particular it explains that the output is simply a static nonlinear function of the input(s) and how membership functions affect the shape of the resulting nonlinearity. Finally there seem to be no examples given in the literature of obtaining fuzzy logic controllers for plant transfer functions which are known to be difficult to control, for example, plants with right half plane zeros, resonant modes or which are unstable. Fuzzy logic ideas have been used in other areas of science and engineering; a text covering fuzzy logic in control is reference 9.22.

## 9.9 Neural Networks

Neural network techniques provide a mechanism for obtaining a static or dynamic nonlinear structure using various types of nonlinear and linear elements, and configurations. Typically parameters in the structure are changed adaptively (training) to try and match given input and corresponding output data structures. Probably the earliest use of this concept in control was a variant on this theme where the objective was to produce a nonlinear function, thus adjusting the system parameters to produce a known output for a given input. This was done to achieve required nonlinear functions on analogue computers where typically the individual functions were a nonlinear element consisting of two adjustable linear segments.

Using sets of input and the corresponding output data, training techniques are used in a neural network to adjust the network coefficients, nonlinearity in the static case, so that they model the system producing the data sets. To test the resulting model it is usual to subject it to new sets of input data to check how closely the resulting output matches the corresponding known output data from the given system. Obviously the success of the approach, as we have seen from the interesting behaviour of nonlinear systems discussed in chapter 1, will be highly dependent on the structure used in the network relative to that of the actual system. This will affect many factors such as the accuracy of the model, the required training time and indeed whether the behaviour can be modelled, for example in the case of chaotic motion.

Although this type of black box approach can be useful, the importance of being able to incorporate physical knowledge into system modelling, particularly the structural connections between linear and nonlinear elements, cannot be over emphasised. The simple example of illustrating the approximation of a static nonlinearity by a neural network provides a good illustration of this. Suppose one has two symmetrical odd nonlinear elements (i) ideal dead zone and (ii) a cubic. Further assume there is a choice of neural network structures for doing the modelling consisting of either (a) linear segmented nonlinearities or (b) power law characteristics. Both of these functions may be described as infinite approximators, as they can approximate any nonlinear function exactly by taking an infinite number of them. Thus if the approximators (b) are used for (i) and (a) for (ii), then indeed to get a good approximation a large number will be required. However, if (a) is used for (i) and (b) for (ii) then only a small finite number will be required to get an exact result. For example, if the power laws are  $x^n$  with  $n = 1, 3, 5$  etc only the cubic term is required for (ii). Neural networks have been used successfully in many areas of science and engineering so there are many books on the topic. Two dealing with their application in control engineering are references 9.23 and 9.24.

## 9.10 Exact Linearization

In section 9.3 the possibility of making a system linear by the addition of other nonlinear elements was discussed from a block diagram viewpoint. Some of the situations mentioned are special cases of an approach, based on state space methods, known as exact linearization. The topic is considered in depth in Isidori [9.25] but introductions are given in several textbooks including, Slotine and Li [9.26], Corriou [9.27] and Glad and Ljung [9.1]. This section will follow quite closely the presentation given in the last reference.

The basic concept of exact linearization is to provide a feedback to the plant input, normally taken as  $u$ , which makes the response from the new input,  $r$ , linear. Consider the state space description

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= n_1(x_1) + n_2(x_2) + u \\ y &= x_1\end{aligned}\tag{9.9}$$

If  $r$  is a reference input then a closed loop control can be taken with

$$u = r - n_1(x_1) - n_2(x_2) - a_1 x_1 - a_2 x_2\tag{9.10}$$

which gives the closed loop system

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= a_1 x_1 - a_2 x_2 + r \\ y &= x_1\end{aligned}\tag{9.11}$$

This is a linear system between the input  $r$  and output  $y$  and the values chosen for  $a_1$  and  $a_2$  allow location of the closed loop poles. When the state space description (9.9) has  $y$  as linear position the technique is often referred to as 'computed force' or when  $y$  is a rotary position as 'computed torque' and is often used in robotics.

On the other hand if the equations are

$$\begin{aligned}\dot{x}_1 &= x_2 + n_1(x_1) \\ \dot{x}_2 &= -x_1 - x_2 + u \\ y &= x_1\end{aligned}\tag{9.12}$$

Then it is not possible to directly linearize  $n_1(x_1)$  by the choice of  $u$  since the two quantities are not in the same equation. Although if the block diagram is drawn it is easily seen how the system can be linearized if  $x_1$  can be measured or estimated using an observer and an additional input given to  $\dot{x}_1$ .  $\dot{x}_1$ , however, is typically an internal variable which cannot be influenced directly. It is easy to see that a general form of state equation which will allow exact feedback linearization if all the states are available, known as the controllability canonical form, with state vector  $\mathbf{x}$ , is :-

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{pmatrix} = \begin{pmatrix} x_2 \\ x_3 \\ \vdots \\ f(\mathbf{x}) + b(\mathbf{x})u \end{pmatrix}, y = x_1 \quad (9.13)$$

with the input  $u$  chosen to be

$$u = (1/b(\mathbf{x}))(r - \mathbf{k}\mathbf{x} - f(\mathbf{x})) \quad (9.14)$$

Here  $u$  consists of three terms, the new input  $r$ , the linear state feedback  $\mathbf{k}\mathbf{x}$  and a term to cancel the nonlinearity. The last line in the state equation becomes

$$\dot{x}_n = r - \mathbf{k}\mathbf{x} \quad (9.15)$$

So the equation is linear between  $r$  and  $y$  and the state feedback gain vector  $\mathbf{k}$  can be chosen to suitably place the system poles. This looks all right in principle but one obvious difficulty is if for some  $\mathbf{x}$ ,  $b(\mathbf{x})$  becomes small or zero. To consider the situation further a more general form of nonlinear state space representation will be considered and some definitions given.

**Try this...**



The sequence 2, 4, 6, 8, 10, 12, 14, 16, ... is the sequence of even whole numbers. The 100th place in this sequence is the number...?

Challenging? Not challenging? Try more >>

[www.alloptions.nl/life](http://www.alloptions.nl/life)

Please click the advert

### 9.10.1 Relative Degree

Consider a system described by the general differential equation

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}) + u\mathbf{g}(\mathbf{x}) \\ y &= h(\mathbf{x})\end{aligned}\tag{9.16}$$

where  $\mathbf{f}$ ,  $\mathbf{g}$  and  $h$  are infinitely differentiable functions of the  $n$ th order state vector  $\mathbf{x}$ . Obviously  $y$  does not depend explicitly on  $u$  so that any changes in  $y$  will occur gradually via  $\mathbf{x}$ .

Differentiating the output equation one obtains

$$\dot{y} = h_x(\mathbf{x})\dot{\mathbf{x}} = h_x(\mathbf{x})(\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})u)\tag{9.17}$$

where  $h_x(\mathbf{x})$  denotes the partial derivative of  $h(\mathbf{x})$  with respect to  $\mathbf{x}$ , that is

$$h_x(\mathbf{x}) = \left( \frac{\partial h}{\partial x_1}, \frac{\partial h}{\partial x_2}, \dots, \frac{\partial h}{\partial x_n} \right)\tag{9.18}$$

Equation (9.17) can be written

$$\dot{y} = h_x\mathbf{f} + uh_x\mathbf{g}\tag{9.19}$$

where the  $\mathbf{x}$  dependence of  $h_x$ ,  $\mathbf{f}$  and  $\mathbf{g}$  has been omitted for simplicity. It can be seen that  $\dot{y}$  now depends directly on  $u$  for at least some values of  $\mathbf{x}$ , if and only if  $h_x\mathbf{g} \neq 0$ , and in this case the system will be said to have relative degree 1. If  $h_x\mathbf{g} \equiv 0$  then one can differentiate again to obtain

$$\ddot{y} = (h_x\mathbf{f})_x\mathbf{f} + u(h_x\mathbf{f})_x\mathbf{g}\tag{9.20}$$

The system is said to have relative degree 2 if  $h_x\mathbf{g} \equiv 0$  and  $(h_x\mathbf{f})_x\mathbf{g} \neq 0$  so that  $\ddot{y}$  depends directly on  $u$ . Otherwise one can continue differentiating until a derivative of  $y$  is directly dependent on  $u$ .

It is useful to introduce some partial differential operators for the repeated differentiations, namely

$$L_f = f_1 \frac{\partial}{\partial x_1} + \dots + f_n \frac{\partial}{\partial x_n} \text{ and } L_g = g_1 \frac{\partial}{\partial x_1} + \dots + g_n \frac{\partial}{\partial x_n}\tag{9.21}$$

$L_f$  is called the Lie derivative in the direction of  $\mathbf{f}$ . If  $v$  denotes the relative degree, which is the lowest derivative of  $y$  that depends on  $u$ , then from the above results it can be seen that

$$\begin{aligned}v = 1: L_g h &\neq 0 \\ v = 2: L_g h &= 0 \quad L_g \mathbf{f} \neq 0 \\ v = 3: L_g L_f h &= 0 \quad L_g L_f^2 h \neq 0\end{aligned}\tag{9.22}$$

etc.

One thus has the following definition

**Definition 1**

The relative degree  $\nu$  is the least positive integer such that  $L_g L_f^{\nu-1} h \neq 0$ .

It can be shown that this definition is consistent with that for a linear system where the relative degree is the difference in the degree between the polynomials of the numerator and denominator terms of the transfer function.

**9.10.2 Input-Output Linearization**

From the above for a system of relative degree  $\nu$ , one has

$$y^{(\nu)} = L_f^\nu h + u L_g L_f^{\nu-1} h \quad (9.23)$$

so that if the feedback, assuming  $L_g L_f^{\nu-1} h \neq 0$ ,

$$u = \frac{1}{L_g L_f^{\nu-1} h} (r - L_f^\nu h) \quad (9.24)$$

is introduced then one has

$$y^{(\nu)} = r \quad (9.25)$$

a linear relation ship between the input  $r$  and the  $\nu$ th derivative of the output. Here  $L_f^\nu h$  and  $L_g L_f^{\nu-1} h$  are functions of  $\mathbf{x}$  and in particular if the latter is zero for any value of the state  $\mathbf{x}$  the required control signal becomes infinite.

If the coordinate change

$$\mathbf{z} = \boldsymbol{\varphi}(\mathbf{x}) \quad (9.26)$$

is made with

$$\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_\nu \\ \phi_{\nu+1} \\ \vdots \\ \phi_n \end{pmatrix} = \begin{pmatrix} h \\ L_f h \\ \vdots \\ L_f^{\nu-1} h \\ \text{arbitrary} \\ \vdots \\ \text{arbitrary} \end{pmatrix} \quad (9.27)$$



Then the new system description will be of the form

$$\begin{pmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \vdots \\ \dot{z}_v \\ \dot{z}_{v+1} \\ \vdots \\ \dot{z}_n \\ y \end{pmatrix} = \begin{pmatrix} z_2 \\ z_3 \\ \vdots \\ \tilde{\xi}(z) + u\eta(z) \\ \psi_1(z, u) \\ \vdots \\ \psi_{n-v}(z, u) \\ z_1 \end{pmatrix} \quad (9.28)$$

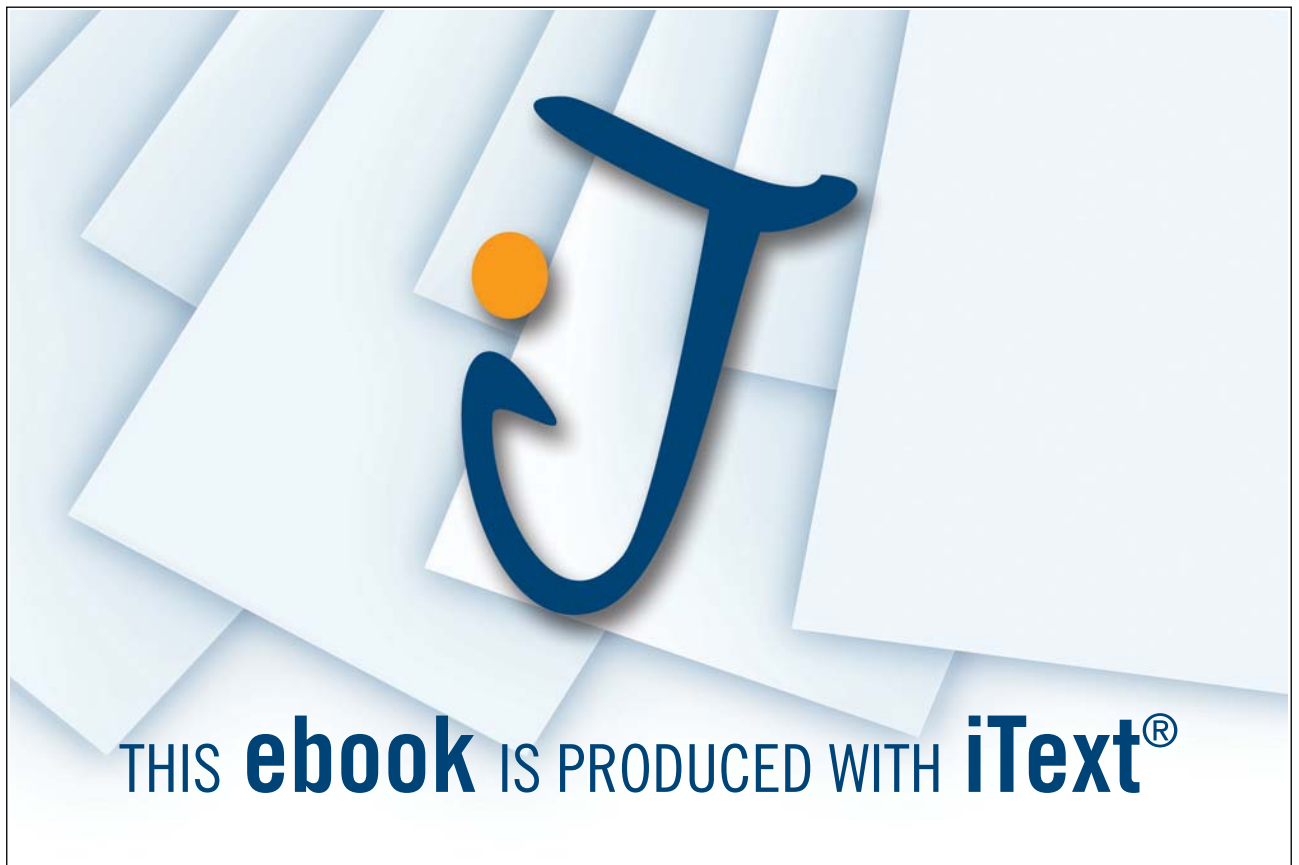
for some functions  $\tilde{\xi}$ ,  $\eta$  and  $\psi$ . The linearizing feedback is

$$u = (r - \tilde{\xi})/\eta \quad (9.29)$$

which gives the closed loop system

$$\begin{pmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \vdots \\ \dot{z}_v \\ \dot{z}_{v+1} \\ \vdots \\ \dot{z}_n \\ y \end{pmatrix} = \begin{pmatrix} z_2 \\ z_3 \\ \vdots \\ r \\ \psi_1(z, (r - \tilde{\xi})/\eta) \\ \vdots \\ \psi_{n-v}(z, (r - \tilde{\xi})/\eta) \\ z_1 \end{pmatrix} \quad (9.30)$$

Please click the advert



Thus the whole system has not been linearised and there is possibly some nonlinear dynamics affecting the states  $z_{v+1}, \dots, z_n$ . This dynamics is not visible in the output and is called the zero dynamics of the system. If the above calculations are performed for a linear system then it is possible to show that the zero dynamics are the zeros of the transfer function from  $u$  to  $y$ .

### 9.10.3 Exact State Linearization

It will be noted from the above that the zero dynamics disappear when  $v = n$ . In this case there is a state feedback which makes the system exactly linear in the state variables

$$z_1 = y, z_2 = \dot{y}, \dots, z_n = y^{(n-1)}$$

and the resultant system will be in the controllability companion form given in equation (9.7) in  $\mathbf{z}$ , that is

$$\begin{pmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \vdots \\ \dot{z}_n \\ y \end{pmatrix} = \begin{pmatrix} z_2 \\ z_3 \\ \vdots \\ \xi(z) + u\eta(z) \\ z_1 \end{pmatrix} \quad (9.31)$$

and can be linearised with the feedback  $u = (r - \xi)/\eta$  and the  $n^{\text{th}}$  line of the above equation is  $\dot{z}_n = r$ .

### 9.10.4 Examples

Two examples are now considered to illustrate the method.

#### Example 1

Consider the nonlinear system

$$\begin{aligned} \dot{x}_1 &= n(x_1) + x_2 \\ \dot{x}_2 &= -x_2 + u \end{aligned}$$

Obviously the input- output linearisation is dependent on the output  $y$  which affects  $h$  in equation (9.27). First let the output be taken as

$$y = x_1 + x_2.$$

Differentiation of  $y$  shows that the first derivative is dependent on  $u$  so that the system is of relative degree 1. From equation (9.27) this means  $y = z_1 = x_1 + x_2$  and  $z_2$  can be chosen arbitrarily. Choosing  $z_2 = x_1$  gives

$$\begin{aligned} \dot{z}_1 &= n(x_1) + u = n(z_2) + u \\ \dot{z}_2 &= \dot{x}_1 = n(x_1) + x_2 = n(z_1) + z_1 - z_2 \end{aligned}$$

Choosing  $u = r - n(z_2)$ , gives  $\dot{z}_1 = r$  and the nonlinear zero dynamics is given by the above equation for  $\dot{z}_2$  and will be stable if  $z_1$  is bounded.

Now as a second choice consider the output to be  $y = x_1$  then differentiating gives

$$\dot{y} = \dot{x}_1 = n(x_1) + x_2 \text{ and}$$

$$\ddot{y} = n'(x_1)\dot{x}_1 + \dot{x}_2 = n'(x_1)(n(x_1) + x_2) - x_2 + u$$

where  $n'(x) = \frac{dn(x)}{dx}$ . Thus choosing  $u = r + x_2 - x_2 n'(x_1) - n'(x_1)n(x_1)$  gives  $\ddot{y} = r$ . In terms of  $z$

$$z_1 = x_1$$

$$z_2 = \dot{x}_1 = n(x_1) + x_2 = \dot{z}_1 \text{ and}$$

$$\dot{z}_2 = r$$

if the feedback  $u$  is taken as

$$u = r - z_2 n'(z_1) + z_2 - n(z_1).$$

More directly from equations (9.24) and (9.27)

$$z_1 = y = x_1 \text{ and}$$

$$z_2 = L_f h = f_1 \frac{\partial h_1}{\partial x_1} + f_2 \frac{\partial h_2}{\partial x_2} = (n_1(x) + x_2) = \dot{x}_1$$

$$L_g L_f h = g_1 \frac{\partial [n(x_1) + x_2]}{\partial x_1} + g_2 \frac{\partial [n(x_1) + x_2]}{\partial x_2} = 0 + 1 = 1 \text{ and}$$

$$L_f^2 h = f_1 \frac{\partial [n(x_1) + x_2]}{\partial x_1} + f_2 \frac{\partial [n(x_1) + x_2]}{\partial x_2} = [n(x_1) + x_2]n'(x_1) - x_2 \text{ giving}$$

$$u = r + x_2 - x_2 n'(x_1) - n'(x_1)n(x_1) \text{ as before.}$$

Consider now that it is desired to place both poles for the closed loop response to  $z_1$  at -1 then choosing

$$u = r - z_1 - 2z_2 - z_2 n'(z_1) + z_2 - n(z_1) \text{ gives}$$

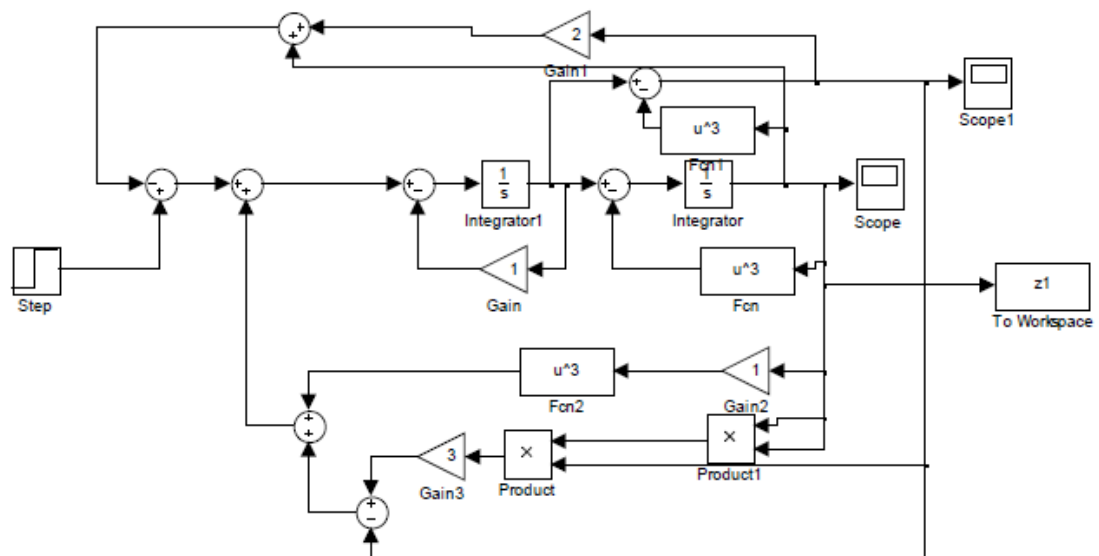
$$\dot{z} = \begin{pmatrix} 0 & 1 \\ -1 & -2 \end{pmatrix} z + \begin{pmatrix} 0 \\ 1 \end{pmatrix} r$$

which has the required characteristic equation of  $s^2 + 2s + 1 = 0$

In particular if  $n(x_1) = -x_1^3$  then the feedback in terms of  $z$  is

$$u = r - z_1 - 2z_2 + 3z_2 z_1^2 + z_2 + z_1^3$$

and Figure 9.10 shows the Simulink diagram for the system. The feedback in the bottom part of the diagram corresponds to the last three terms in the above equation for  $u$ , which without further terms would make  $\dot{z}_2 = r$ . There are two positive feedback nonlinear terms in these three to make the system linear between the input  $r$  and  $z$ , and the other two terms are the linear ones to place both poles at -1. The linear responses to  $z_1$  (Scope) and  $z_2$  (Scope1) for a step input as expected correspond to those of transfer functions  $1/(s+1)^2$  and  $s/(s+1)^2$ , respectively and that to  $z_1$  is shown in Figure 9.11.



**Figure 9.10** Simulink diagram for example 1



# The next step for top-performing graduates

## Masters in Management

Designed for high-achieving graduates across all disciplines, London Business School's Masters in Management provides specific and tangible foundations for a successful career in business.

This 12-month, full-time programme is a business qualification with impact. In 2010, our MiM employment rate was 95% within 3 months of graduation\*; the majority of graduates choosing to work in consulting or financial services.

As well as a renowned qualification from a world-class business school, you also gain access to the School's network of more than 34,000 global alumni – a community that offers support and opportunities throughout your career.

For more information visit [www.london.edu/mm](http://www.london.edu/mm), email [mim@london.edu](mailto:mim@london.edu) or give us a call on **+44 (0)20 7000 7573**.

\* Figures taken from London Business School's Masters in Management 2010 employment report

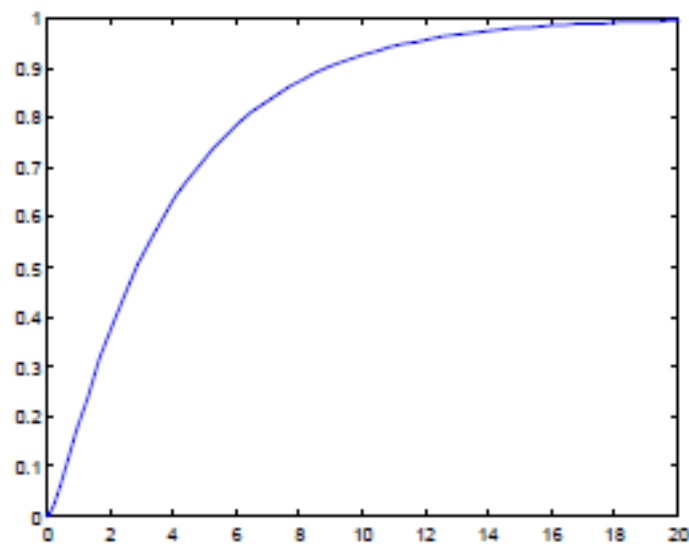


Figure 9.11 Input-output step response for example 1.

This appears to be a rather complicated but interesting way of obtaining a linear input-output response but what happens if there are modelling errors? To illustrate the problems for this particular case let the actual nonlinearity be  $n(x_1) = -ax_1^3$  with 10% variations in  $a$ , that is with  $a$  equal to first 0.9 and then 1.1. Step responses are shown in Figure 9.12 for step magnitudes of 1 and 1.14, respectively. The four responses are marked in the boxes with the first number being,  $a$ , and the second the step magnitude. Although three responses are satisfactory the fourth for a step input magnitude of 1.14 with  $a = 0.9$  actually becomes unstable. This is not surprising in view of the strong positive feedback, whereas the stability of the original system is not a problem.

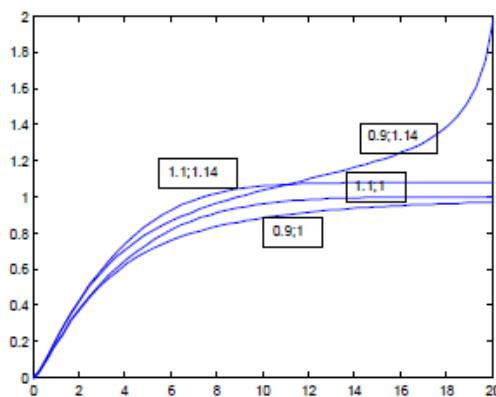


Figure 9.12 Step responses for Example 1 with nonlinearity parameter changes.

**Example 2**

Consider the third order system

$$\begin{aligned}\dot{x}_1 &= x_2 - ax_1 \\ \dot{x}_2 &= -bx_2 + n(x_3) \\ \dot{x}_3 &= -cx_3 + u \\ y &= x_1\end{aligned}$$

where  $n(x_3)$  is any nonlinear element.

This means

$$h = x_1 \quad \mathbf{f} = \begin{pmatrix} x_2 - ax_1 \\ -bx_2 + n(x_3) \\ -cx_3 \end{pmatrix} \text{ and } \mathbf{g} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \text{ so that}$$

$$\begin{aligned}L_f h &= f_1 = x_2 - ax_1 \\ L_f^2 h &= f_1(-a) + f_2 = -ax_2 + a^2 x_1 - bx_2 + n(x_3) \\ L_f^3 h &= f_1(a^2) + f_2(-(a+b)) + f_3 n(x_3) \\ &= -a^3 x_1 + (a^2 + ab + b^2)x_2 - (a+b)n(x_3) - cx_3 n'(x_3)\end{aligned}$$

and

$$L_g L_f^2 h = n'(x_3)$$

Thus the required feedback to make  $\dot{z}_n = r$  is

$$u = \frac{1}{n'(x_3)} \{r + a^3 x_1 - (a^2 + ab + b^2)x_2 + (a+b)n(x_3) + cx_3 n'(x_3)\}$$

or in terms of  $\mathbf{z}$  apart from  $n'(x_3)$

$$u = \frac{1}{n'(x_3)} \{r + abz_2 + (a+b)z_3 + cx_3 n'(x_3)\}$$

Thus if the three poles for the linear system in  $\mathbf{z}$  are all to be placed at -1 then the required  $u$  is given by

$$u = \frac{1}{n'(x_3)} \{r + abz_2 + (a+b)z_3 + cx_3 n'(x_3) - z_1 - 3z_2 - 3z_3\}.$$

Here if  $n(x_3) = x_3^3$  then  $n'(x_3)$  needs to be limited to a small value when  $x_3 = 0$  to limit the value of the signal  $u$ . It is also clear that the method can easily be used if the nonlinearity is linear segmented as its slope is known. Figure 9.13 shows the implementation in Simulink for the case of  $a = 0$  and  $b = c = 1$  and the linear segmented nonlinearity

$$\begin{aligned}n(x_3) &= x_3 & \text{for } |x_3| < 1 \\ &= 0.3x_3 + 0.7 & \text{for } x_3 > 1 \\ &= 0.3x_3 - 0.7 & \text{for } x_3 < -1\end{aligned}$$

The Scope blocks are connected to the components of  $\mathbf{z}$  and the output of the product block is the signal  $u$ . The relay elements connected in parallel are used to give  $n'(x_3)$ .

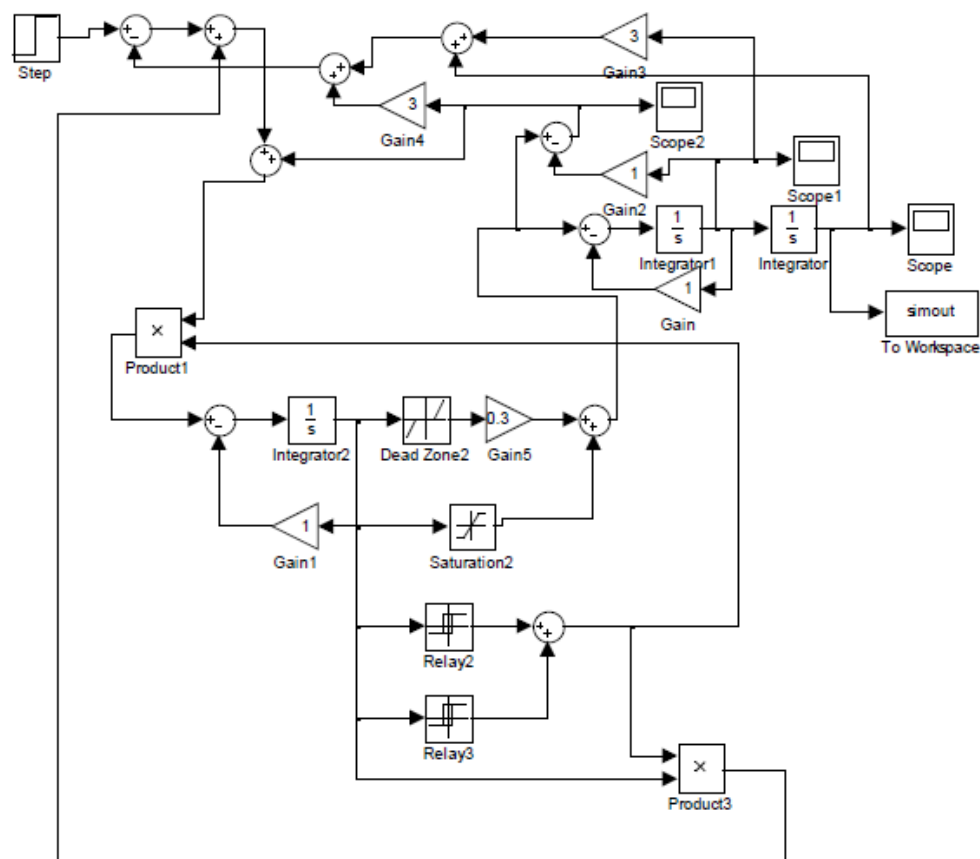


Figure 9.13 Simulink implementation for Example 2

### 9.10.5 Further Comments

To summarise, the procedure described in the previous sections for a system with relative degree,  $\nu$ , equal to the dimension of the state  $\mathbf{x}$  allows a nonlinear transformation to be found from  $\mathbf{x}$  to a new state  $\mathbf{z}$  and a control signal  $u$  which will make  $\mathbf{z}$  linear with respect to the new input  $r$ . A feedback controller can then be designed in terms of the original state  $\mathbf{x}$ , or  $\mathbf{z}$ , to control the linear dynamics of  $\mathbf{z}$  with respect to  $r$ . If the relative degree,  $\nu$ , is less than  $n$  the dynamics of some states of  $\mathbf{z}$ , known as the zero dynamics will remain nonlinear and need to be checked in case they are unstable. The required control signal  $u$  is given in equation (9.24) and since its denominator is a function of the state  $\mathbf{x}$  it may not be finite at all points in the state space. The method requires the state  $\mathbf{x}$  to be available so these values must either be measured or estimated using a nonlinear observer.

Perhaps the major advantage of the method is that it allows a stable design of a nonlinear system but requires an accurate mathematical model. When the assumed model does not accurately match the system it may be difficult to assess the effects it may have and as shown by the first example the effects may be undesirable. Also as mentioned previously it may not be appropriate to remove the effects of the nonlinearity, in this case by using control energy, as their existence may not be detrimental to the system performance.



## 9.11 General Comments

Several methods for the design of controllers for nonlinear systems have been briefly covered in this chapter. There is a huge literature on many of the topics so a few appropriate references have been cited but thousands more can be found by putting topics and key words mentioned in the above into Google. Comparison of the various approaches is not easy except perhaps in the case of optimal control where a solution is obtained which meets a certain criterion, and therefore by definition is optimal for that case. However, even in this case, the criterion may have had to be perturbed from the true requirements in order to obtain a mathematical solution or, alternatively, it might be known that the model used has deficiencies. A large number of published papers are usually devoted to a particular approach not to comparisons of methods whilst in practice the approach used will be influenced by many factors not all of which will be easy to describe analytically.

Finally, one must again stress the importance of simulation to support studies of nonlinear control systems no matter what analytical theories are used.

## 9.12 References

- 9.1 Glad, T & Ljung, L 2000, *Control Theory: Multivariable and Nonlinear Methods*, Taylor and Francis, London.
- 9.2 Friedland, B 1996, *Advanced Control System Design*, Prentice Hall, New Jersey.
- 9.3 Rugh, WJ 1991, Analytic Framework for Gain-Scheduling, *IEEE Control Systems Magazine*, Vol 11, pp 79-84.
- 9.4 Rugh, WJ & Shamma, JS 2000, Research on gain scheduling, *Automatica*, Vol 36, pp 1401-1425.
- 9.5 Murray-Smith, R & Johansen, TA, (Editors) 1997, *Multiple Model Approaches to Modelling and Control*, Taylor and Francis, London.
- 9.6 Bohn, C & Atherton, DP 1995, An Analysis Package Comparing PID Anti-Windup Strategies. *IEEE Control Systems Magazine*, Vol 15, pp 34-40.
- 9.7 Nanka-Bruce, O and Atherton, DP 1990, Design of Nonlinear Controllers for Nonlinear Plants, *IFAC Congress, Tallinn*, Vol 6, pp 75-80.
- 9.8 Atherton, DP, Benouarets, M & Nanka-Bruce, O 1993, Design of Nonlinear PID Controllers for Nonlinear Plants. *Proc. IFAC World Congress, Sydney*, Vol 3, pp 355-358.
- 9.9 McCausland, I 1969, *Introduction to Optimal Control*, Wiley, New York.
- 9.10 Jacobs, OLR 1974, *Introduction to Control Theory*, Chapter 9, Oxford University Press.
- 9.11 Ryan, EP 1982, *Optimal Relay and Saturating Control System Synthesis*, Peter Peregrinus Ltd. Stevenage.
- 9.12 Athans, M & Falb, PL 1966, *Optimal Control*. McGraw-Hill, New York.
- 9.13 Lewis, FL & Syrmos, VL 1995, *Optimal Control*, John Wiley, New York.
- 9.14 West, JC 1960, *Analytical Techniques for Nonlinear Control Systems*, Chapter 6, EUP London.
- 9.15 Emelyanov, SV 1967, *Variable Structure Control*, Nauka, Moscow.
- 9.16 Utkin, VI 1992, *Sliding Modes in Control Optimization*, Springer-Verlag, Berlin.
- 9.17 Zinober, ASI (Editor) 1990, *Deterministic Control of Uncertain Systems*, Peter Peregrinus, Stevenage.
- 9.18 Edwards, C & Spurgeon, SK 1998, *Sliding Mode Control*, Taylor and Francis, London.
- 9.19 Zadeh, LA 1965, Fuzzy Sets, *Information and Control*, Vol 8, pp 338-353.
- 9.20 Mamdani, E H & Assilian, S 1975, Experiment in Linguistic Synthesis with A Fuzzy Logic Controller, *International Journal of Man-Machine Studies*, vol. 7, pp 1-13.

- 9.21 Schmid, Chr. 2004, Dynlab course at <http://virtual.cvut.cz/dynlabmods/syscontrol.pdf>. Chapters 14-17.
- 9.22 Verbruggen, HB & Babuska, R (Editors) 1999, Fuzzy Logic Control; Advances in Applications. World Scientific Publishing Co.
- 9.23 Warwick, K, Irwin, GW & Hunt, KJ (Editors) 1992, Neural Networks for Control and Systems, Peter Peregrinus, Stevenage.
- 9.24 Irwin, GW, Warwick, K & Hunt KJ (Editors) 1995, Neural Network Applications in Control, Peter Perigrinus, Stevenage.
- 9.25. Isidori, A 1995, Nonlinear Control Systems; An Introduction, 3rd Edition, Springer Verlag, New York.
- 9.26 Slotine, J-J & Li, W 1990, Applied Nonlinear Control, Prentice Hall, New Jersey.
- 9.27 Corriou, J-P 2004, Process Control; Theory and Applications, Chapter 17, Springer Verlag London.

Please click the advert



You're full of *energy*  
*and ideas*. And that's  
just what we are looking for.

Looking for a career where your ideas could really make a difference? UBS's Graduate Programme and internships are a chance for you to experience for yourself what it's like to be part of a global team that rewards your input and believes in succeeding together.

Wherever you are in your academic career, make your future a part of ours by visiting [www.ubs.com/graduates](http://www.ubs.com/graduates).

[www.ubs.com/graduates](http://www.ubs.com/graduates)



© UBS 2010. All rights reserved.