Applied Nonlinear Control

JEAN-JACQUES E. SLOTINE
Massachusetts Institute of Technology

WEIPING LI
Massachusetts Institute of Technology

Prentice Hall
Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging-in-Publication Data

Slotine, J.-J. E. (Jean-Jacques E.)
Applied nonlinear control / Jean-Jacques E. Slotine, Weiping Li

p. cm.
Includes bibliographical references.
ISBN 0-13-040890-5
1, Nonlinear control theory. I. Li, Weiping. II. Title.
QA402.35.S56 1991 90-33365
629.8'312—dc20 CIP

Editorial/production supervision and
interior design: JENNIFER WENZEL
Cover design: KAREN STEPHENS
Manufacturing Buyer: LORI BULWIN

Printed in the United States of America

10 9 8 7 6

ISBN 0-13-040890-5

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

To Our Parents

Contents

Preface

In recent years, the availability of powerful low-cost microprocessors has spurred great advances in the theory and applications of nonlinear control. In terms of theory, major strides have been made in the areas of feedback linearization, sliding control, and nonlinear adaptation techniques. In terms of applications, many practical nonlinear control systems have been developed, ranging from digital "fly-by-wire" flight control systems for aircraft, to "drive-by-wire" automobiles, to advanced robotic and space systems. As a result, the subject of nonlinear control is occupying an increasingly important place in automatic control engineering, and has become a necessary part of the fundamental background of control engineers.

This book, based on a course developed at MIT, is intended as a textbook for senior and graduate students, and as a self-study book for practicing engineers. Its objective is to present the fundamental results of modern nonlinear control while keeping the mathematical complexity to a minimum, and to demonstrate their use and implications in the design of practical nonlinear control systems. Although a major motivation of this book is to detail the many recent developments in nonlinear control, classical techniques such as phase plane analysis and the describing function method are also treated, because of their continued practical importance.

In order to achieve our fundamental objective, we have tried to bring the following features to this book:

- **Readability:** Particular attention is paid to the readability of the book by carefully organizing the concepts, intuitively interpreting the major results, and selectively using the mathematical tools. The readers are only assumed to have had one introductory control course. No mathematical background beyond ordinary differential equations and elementary matrix algebra is required. For each new result, interpretation is emphasized rather than mathematics. For each major result, we try to ask and answer the following key questions: What does the result intuitively and physically mean? How can it be applied to practical problems? What is its relationship to other theorems? All major concepts and results are demonstrated by examples. We believe that learning and generalization from examples are crucial for proficiency in applying any theoretical result.

- **Practicality:** The choice and emphasis of materials is guided by the basic objective of making an engineer or student capable of dealing with practical control problems in industry. Some results of mostly theoretical interest are not included. The selected materials, in one way or another, are intended to allow readers to gain insights into the solution of real problems.

• **Comprehensiveness:** The book contains both classical materials, such as Lyapunov analysis and describing function techniques, and more modern topics such as feedback linearization, adaptive control, and sliding control. To facilitate digestion, asterisks are used to indicate sections which, given their relative complexity, can be safely skipped in a first reading.

• **Currentness:** In the past few years, a number of major results have been obtained in nonlinear control, particularly in nonlinear control system design and in robotics. It is one of the objectives of this book to present these new and important developments, and their implications, in a clear, easily understandable fashion. The book can thus be used as a reference and a guide to the active literature in these fields.

The book is divided into two major parts. Chapters 2-5 present the major *analytical* tools that can be used to study a nonlinear system, while chapters 6-9 treat the major nonlinear controller *design* techniques. Each chapter is supplied with exercises, allowing the reader to further explore specific aspects of the material discussed. A detailed index and a bibliography are provided at the end of the book.

The material included exceeds what can be taught in one semester or self-learned in a short period. The book can be studied in many ways, according to the particular interests of the reader or the instructor. We recommend that a first reading include a detailed study of chapter 3 (basic Lyapunov theory), sections 4.5-4.7 (Barbalat's lemma and passivity tools), section 6.1 and parts of sections 6.2-6.4 (feedback linearization), chapter 7 (sliding control), sections 8.1-8.3 and 8.5 (adaptive control of linear and nonlinear systems), and chapter 9 (control of multi-input physical systems). Conversely, sections denoted with an asterisk can be skipped in a first reading.

Many colleagues, students, and friends greatly contributed to this book through stimulating discussions and judicious suggestions. Karl Hedrick provided us with continued enthusiasm and encouragement, and with many valuable comments and suggestions. Discussions with Karl Aström and Semyon Meerkov helped us better define the tone of the book and its mathematical level. Harry Asada, Jo Bentsman, Marika DiBenedetto, Olav Egeland, Neville Hogan, Marija Ilic, Lars Nielsen, Ken Salisbury, Sajhendra Singh, Mark Spong, David Wormley, and Dana Yoerger provided many useful suggestions and much moral support. Barbara Hove created most of the nicer drawings in the book; Günter Niemeyer's expertise and energy was invaluable in setting up the computing and word processing environments; Hyun Yang greatly helped with the computer simulations; all three provided us with extensive technical and editorial comments. The book also greatly benefited from the interest and enthusiasm of many students who took the course at MIT.

Partial summer support for the first author towards the development of the book was provided by Gordon Funds. Finally, the energy and professionalism of Tim Bozik and Jennifer Wenzel at Prentice-Hall were very effective and highly appreciated.

Jean-Jacques E. Slotine
Weiping Li

Applied Nonlinear Control

Chapter 1
Introduction

The subject of nonlinear control deals with the analysis and the design of nonlinear control systems, *i.e.*, of control systems containing at least one nonlinear component. In the analysis, a nonlinear closed-loop system is assumed to have been designed, and we wish to determine the characteristics of the system's behavior. In the design, we are given a nonlinear plant to be controlled and some specifications of closed-loop system behavior, and our task is to construct a controller so that the closed loop system meets the desired characteristics. In practice, of course, the issues of design and analysis are intertwined, because the design of a nonlinear control system usually involves an iterative process of analysis and design.

This introductory chapter provides the background for the specific analysis and design methods to be discussed in the later chapters. Section 1.1 explains the motivations for embarking on a study of nonlinear control. The unique and rich behaviors exhibited by nonlinear systems are discussed in section 1.2. Finally, section 1.3 gives an overview of the organization of the book.

1.1 Why Nonlinear Control ?

Linear control is a mature subject with a variety of powerful methods and a long history of successful industrial applications. Thus, it is natural for one to wonder why so many researchers and designers, from such broad areas as aircraft and spacecraft control, robotics, process control, and biomedical engineering, have recently showed an active interest in the development and applications of nonlinear control methodologies. Many reasons can be cited for this interest:

• **Improvement of existing control systems**: Linear control methods rely on the key assumption of small range operation for the linear model to be valid. When the required operation range is large, a linear controller is likely to perform very poorly or to be unstable, because the nonlinearities in the system cannot be properly compensated for. Nonlinear controllers, on the other hand, may handle the nonlinearities in large range operation directly. This point is easily demonstrated in robot motion control problems. When a linear controller is used to control robot motion, it neglects the nonlinear forces associated with the motion of the robot links. The controller's accuracy thus quickly degrades as the speed of motion increases, because many of the dynamic forces involved, such as Coriolis and centripetal forces, vary as the square of the speed. Therefore, in order to achieve a pre-specified accuracy in robot tasks such as pick-and-place, arc welding and laser cutting, the speed of robot motion, and thus productivity, has to be kept low. On the other hand, a conceptually simple nonlinear controller, commonly called computed torque controller, can fully compensate the nonlinear forces in the robot motion and lead to high accuracy control for a very large range of robot speeds and a large workspace.

• **Analysis of hard nonlinearities:** Another assumption of linear control is that the system model is indeed linearizable. However, in control systems there are many nonlinearities whose discontinuous nature does not allow linear approximation. These so-called "hard nonlinearities" include Coulomb friction, saturation, dead-zones, backlash, and hysteresis, and are often found in control engineering. Their effects cannot be derived from linear methods, and nonlinear analysis techniques must be developed to predict a system's performance in the presence of these inherent nonlinearities. Because such nonlinearities frequently cause undesirable behavior of the control systems, such as instabilities or spurious limit cycles, their effects must be predicted and properly compensated for.

• **Dealing with model uncertainties:** In designing linear controllers, it is usually necessary to assume that the parameters of the system model are reasonably well known. However, many control problems involve uncertainties in the model parameters. This may be due to a slow time variation of the parameters (*e.g.*, of ambient air pressure during an aircraft flight), or to an abrupt change in parameters (*e.g,*, in the inertial parameters of a robot when a new object is grasped). A linear controller based on inaccurate or obsolete values of the model parameters may exhibit significant performance degradation or even instability. Nonlinearities can be intentionally introduced into the controller part of a control system so that model uncertainties can be tolerated. Two classes of nonlinear controllers for this purpose are robust controllers and adaptive controllers.

• **Design Simplicity:** Good nonlinear control designs may be simpler and more intuitive than their linear counterparts. This *a priori* paradoxical result comes from the fact that nonlinear controller designs are often deeply rooted in the physics of the plants. To take a very simple example, consider a swinging pendulum attached to a hinge, in the vertical plane. Starting from some arbitrary initial angle, the pendulum will oscillate and progressively stop along the vertical. Although the pendulum's behavior could be analyzed close to equilibrium by linearizing the system, physically its stability has very little to do with the eigenvalues of some linearized system matrix: it comes from the fact that the total mechanical energy of the system is progressively dissipated by various friction forces (*e.g.*, at the hinge), so that the pendulum comes to rest at a position of minimal energy.

There may be other related or unrelated reasons to use nonlinear control techniques, such as cost and performance optimality. In industrial settings, ad-hoc extensions of linear techniques to control advanced machines with significant nonlinearities may result in unduly costly and lengthy development periods, where the control code comes with little stability or performance guarantees and is extremely hard to transport to similar but different applications. Linear control may require high quality actuators and sensors to produce linear behavior in the specified operation range, while nonlinear control may permit the use of less expensive components with nonlinear characteristics. As for performance optimality, we can cite bang-bang type controllers, which can produce fast response, but are inherently nonlinear.

Thus, the subject of nonlinear control is an important area of automatic control. Learning basic techniques of nonlinear control analysis and design can significantly enhance the ability of a control engineer to deal with practical control problems effectively. It also provides a sharper understanding of the real world, which is inherently nonlinear. In the past, the application of nonlinear control methods had been limited by the computational difficulty associated with nonlinear control design and analysis. In recent years, however, advances in computer technology have greatly relieved this problem. Therefore, there is currently considerable enthusiasm for the research and application of nonlinear control methods. The topic of nonlinear control design for large range operation has attracted particular attention because, on the one hand, the advent of powerful microprocessors has made the implementation of nonlinear controllers a relatively simple matter, and, on the other hand, modern technology, such as high-speed high-accuracy robots or high-performance aircrafts, is demanding control systems with much more stringent design specifications. Nonlinear control occupies an increasingly conspicuous position in control engineering, as reflected by the ever-increasing number of papers and reports on nonlinear control research and applications.

1.2 Nonlinear System Behavior

Physical systems are inherently nonlinear. Thus, all control systems are nonlinear to a certain extent. Nonlinear control systems can be described by nonlinear differential equations. However, if the operating range of a control system is small, and if the involved nonlinearities are smooth, then the control system may be reasonably approximated by a linearized system, whose dynamics is described by a set of linear differential equations.

NONLINEARITIES

Nonlinearities can be classified as *inherent (natural)* and *intentional (artificial)*. Inherent nonlinearities are those which naturally come with the system's hardware and motion. Examples of inherent nonlinearities include centripetal forces in rotational motion, and Coulomb friction between contacting surfaces. Usually, such nonlinearities have undesirable effects, and control systems have to properly compensate for them. Intentional nonlinearities, on the other hand, are artificially introduced by the designer. Nonlinear control laws, such as adaptive control laws and bang-bang optimal control laws, are typical examples of intentional nonlinearities.

Nonlinearities can also be classified in terms of their mathematical properties, as *continuous* and *discontinuous*. Because discontinuous nonlinearities cannot be locally approximated by linear functions, they are also called "hard" nonlinearities. Hard nonlinearities (such as, *e.g.*, backlash, hysteresis, or stiction) are commonly found in control systems, both in small range operation and large range operation. Whether a system in small range operation should be regarded as nonlinear or linear depends on the magnitude of the hard nonlinearities and on the extent of their effects on the system performance. A detailed discussion of hard nonlinearities is provided in section 5.2.

LINEAR SYSTEMS

Linear control theory has been predominantly concerned with the study of linear time-invariant (LTI) control systems, of the form

$$\dot{\mathbf{x}} = \mathbf{A}\,\mathbf{x} \qquad (1.1)$$

with $\mathbf{x}$ being a vector of states and $\mathbf{A}$ being the system matrix. LTI systems have quite simple properties, such as

• a linear system has a *unique equilibrium point* if **A** is nonsingular;

• the equilibrium point is stable if all eigenvalues of **A** have negative real parts, *regardless of initial conditions*;

• the transient response of a linear system is composed of the natural modes of the system, and the general solution can be solved analytically;

• in the presence of an external input $\mathbf{u}(t)$, *i.e.*, with

$$\dot{\mathbf{x}} = \mathbf{A}\,\mathbf{x} + \mathbf{B}\,\mathbf{u} \tag{1.2}$$

the system response has a number of interesting properties. First, it satisfies the *principle of superposition*. Second, the asymptotic stability of the system (1.1) implies bounded-input bounded-output stability in the presence of **u**. Third, a sinusoidal input leads to a sinusoidal output of the same frequency.

AN EXAMPLE OF NONLINEAR SYSTEM BEHAVIOR

The behavior of nonlinear systems, however, is much more complex. Due to the lack of linearity and of the associated superposition property, nonlinear systems respond to external inputs quite differently from linear systems, as the following example illustrates.

Example 1.1: A simplified model of the motion of an underwater vehicle can be written

$$\dot{v} + |v|\, v = u \tag{1.3}$$

where v is the vehicle velocity and u is the control input (the thrust provided by a propeller). The nonlinearity $|v|\, v$ corresponds to a typical "square-law" drag.

Assume that we apply a unit step input in thrust u, followed 5 seconds later by a negative unit step input. The system response is plotted in Figure 1.1. We see that the system settles much faster in response to the positive unit step than it does in response to the subsequent negative unit step. Intuitively, this can be interpreted as reflecting the fact that the "apparent damping" coefficient $|v|$ is larger at high speeds than at low speeds.

Assume now that we repeat the same experiment but with larger steps, of amplitude 10. Predictably, the difference between the settling times in response to the positive and negative steps is even more marked (Figure 1.2). Furthermore, the settling speed v_s in response to the first step is *not* 10 times that obtained in response to the first unit step in the first experiment, as it would be in a linear system. This can again be understood intuitively, by writing that

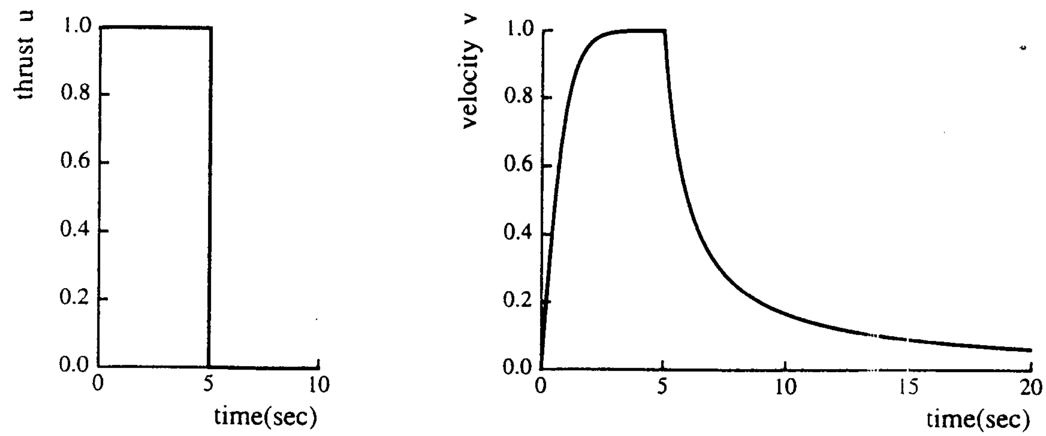

Figure 1.1 : Response of system (1.3) to unit steps

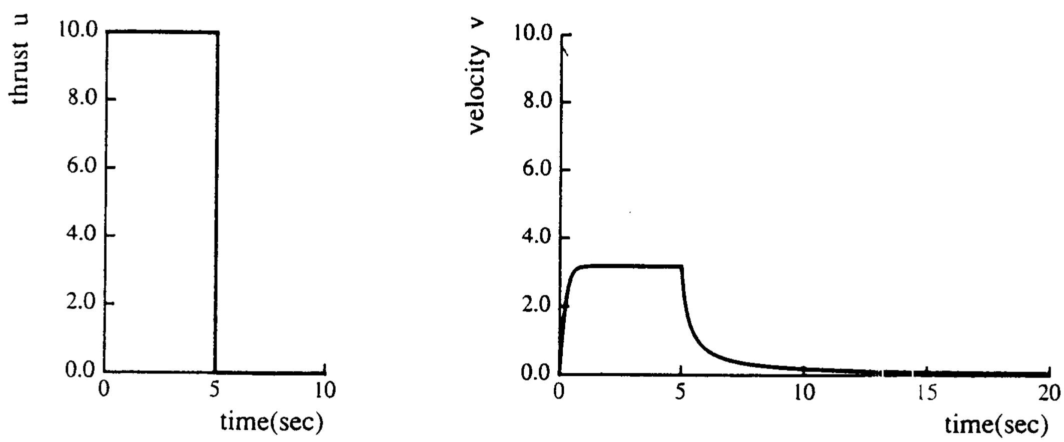

Figure 1.2 : Response of system (1.3) to steps of amplitude 10

$$u = 1 \quad \Rightarrow \quad 0 + |v_s|\, v_s = 1 \quad \Rightarrow \quad v_s = 1$$

$$u = 10 \quad \Rightarrow \quad 0 + |v_s|\, v_s = 10 \quad \Rightarrow \quad v_s = \sqrt{10} \approx 3.2$$

Carefully understanding and effectively controlling this nonlinear behavior is particularly important if the vehicle is to move in a large dynamic range and change speeds continually, as is typical of industrial remotely-operated underwater vehicles (R.O.V.'s). □

SOME COMMON NONLINEAR SYSTEM BEHAVIORS

Let us now discuss some common nonlinear system properties, so as to familiarize ourselves with the complex behavior of nonlinear systems and provide a useful background for our study in the rest of the book.

Multiple Equilibrium Points

Nonlinear systems frequently have more than one equilibrium point (an equilibrium point is a point where the system can stay forever without moving, as we shall formalize later). This can be seen by the following simple example.

Example 1.2: A first-order system

Consider the first order system

$$\dot{x} = -x + x^2 \tag{1.4}$$

with initial condition $x(0) = x_o$. Its linearization is

$$\dot{x} = -x \tag{1.5}$$

The solution of this linear equation is $x(t) = x_o\, e^{-t}$. It is plotted in Figure 1.3(a) for various initial conditions. The linearized system clearly has a unique equilibrium point at $x = 0$.

By contrast, integrating equation $dx/(-x + x^2) = dt$, the actual response of the nonlinear dynamics (1.4) can be found to be

$$x(t) = \frac{x_o\, e^{-t}}{1 - x_o + x_o e^{-t}}$$

This response is plotted in Figure 1.3(b) for various initial conditions. The system has two equilibrium points, $x = 0$ and $x = 1$, and its qualitative behavior strongly depends on its initial condition. □

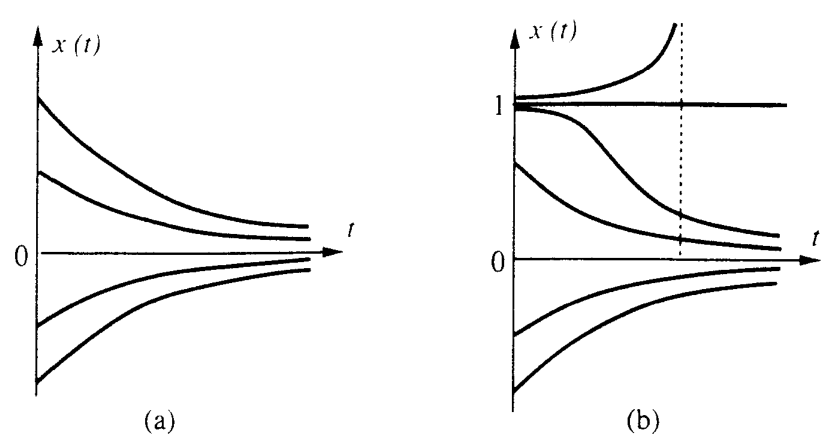

Figure 1.3 : Responses of the linearized system (a) and the nonlinear system (b)

The issue of motion stability can also be discussed with the aid of the above example. For the linearized system, stability is seen by noting that for *any* initial condition, the motion always converges to the equilibrium point $x = 0$. However, consider now the actual nonlinear system. While motions starting with $x_o < 1$ will indeed converge to the equilibrium point $x = 0$, those starting with $x_o > 1$ will go to infinity (actually in finite time, a phenomenon known as finite escape time). This means that the stability of nonlinear systems may depend on initial conditions.

In the presence of a bounded external input, stability may also be dependent on the input value. This input dependence is highlighted by the so-called bilinear system

$$\dot{x} = x u$$

If the input u is chosen to be -1, then the state x converges to 0. If $u = 1$, then $|x|$ tends to infinity.

Limit Cycles

Nonlinear systems can display oscillations of fixed amplitude and fixed period without external excitation. These oscillations are called limit cycles, or self-excited oscillations. This important phenomenon can be simply illustrated by a famous oscillator dynamics, first studied in the 1920's by the Dutch electrical engineer Balthasar Van der Pol.

Example 1.3: Van der Pol Equation

The second-order nonlinear differential equation

$$m\ddot{x} + 2c(x^2 - 1)\dot{x} + kx = 0 \tag{1.6}$$

where m, c and k are positive constants, is the famous Van der Pol equation. It can be regarded as describing a mass-spring-damper system with a position-dependent damping coefficient $2c(x^2 - 1)$ (or, equivalently, an RLC electrical circuit with a nonlinear resistor). For large values of x, the damping coefficient is positive and the damper removes energy from the system. This implies that the system motion has a convergent tendency. However, for small values of x, the damping coefficient is negative and the damper adds energy into the system. This suggests that the system motion has a divergent tendency. Therefore, because the nonlinear damping varies with x, the system motion can neither grow unboundedly nor decay to zero. Instead, it displays a sustained oscillation independent of initial conditions, as illustrated in Figure 1.4. This so-called limit cycle is sustained by periodically releasing energy into and absorbing energy from the environment, through the damping term. This is in contrast with the case of a conservative mass-spring system, which does not exchange energy with its environment during its vibration. □

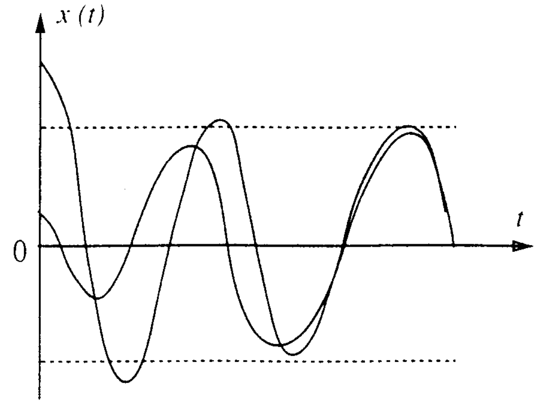

Figure 1.4 : Responses of the Van der Pol oscillator

Of course, sustained oscillations can also be found in linear systems, in the case of marginally stable linear systems (such as a mass-spring system without damping) or in the response to sinusoidal inputs. However, limit cycles in nonlinear systems are different from linear oscillations in a number of fundamental aspects. First, the amplitude of the self-sustained excitation is independent of the initial condition, as seen in Figure 1.2, while the oscillation of a marginally stable linear system has its amplitude determined by its initial conditions. Second, marginally stable linear systems are very sensitive to changes in system parameters (with a slight change capable of leading either to stable convergence or to instability), while limit cycles are not easily affected by parameter changes.

Limit cycles represent an important phenomenon in nonlinear systems. They can be found in many areas of enginering and nature. Aircraft wing fluttering, a limit cycle caused by the interaction of aerodynamic forces and structural vibrations, is frequently encountered and is sometimes dangerous. The hopping motion of a legged robot is another instance of a limit cycle. Limit cycles also occur in electrical circuits, *e.g.*, in laboratory electronic oscillators. As one can see from these examples, limit cycles can be undesirable in some cases, but desirable in other cases. An engineer has to know how to eliminate them when they are undesirable, and conversely how to generate or amplify them when they are desirable. To do this, however, requires an understanding of the properties of limit cycles and a familiarity with the tools for manipulating them.

Bifurcations

As the parameters of nonlinear dynamic systems are changed, the stability of the equilibrium point can change (as it does in linear systems) and so can the number of equilibrium points. Values of these parameters at which the qualitative nature of the system's motion changes are known as *critical* or *bifurcation* values. The phenomenon of bifurcation, *i.e.*, quantitative change of parameters leading to qualitative change of system properties, is the topic of bifurcation theory.

For instance, the smoke rising from an incense stick (smokestacks and cigarettes are old-fashioned) first accelerates upwards (because it is lighter than the ambient air), but beyond some critical velocity breaks into swirls. More prosaically, let us consider the system described by the so-called undamped Duffing equation

$$\ddot{x} + \alpha x + x^3 = 0$$

(the damped Duffing equation is $\ddot{x} + c\dot{x} + \alpha x + \beta x^3 = 0$, which may represent a mass-damper-spring system with a hardening spring). We can plot the equilibrium points as a function of the parameter α. As α varies from positive to negative, one equilibrium point splits into *three* points ($x_e = 0, \sqrt{\alpha}, -\sqrt{\alpha}$), as shown in Figure 1.5(a). This represents a qualitative change in the dynamics and thus $\alpha = 0$ is a critical bifurcation value. This kind for bifurcation is known as a *pitchfork*, due to the shape of the equilibrium point plot in Figure 1.5(a).

Another kind of bifurcation involves the emergence of limit cycles as parameters are changed. In this case, a pair of complex conjugate eigenvalues $p_1 = \gamma + j\omega$, $p_2 = \gamma - j\omega$ cross from the left-half plane into the right-half plane, and the response of the unstable system diverges to a limit cycle. Figure 1.5(b) depicts the change of typical system state trajectories (states are x and $\dot{x}$) as the parameter α is varied. This type of bifurcation is called a Hopf bifurcation.

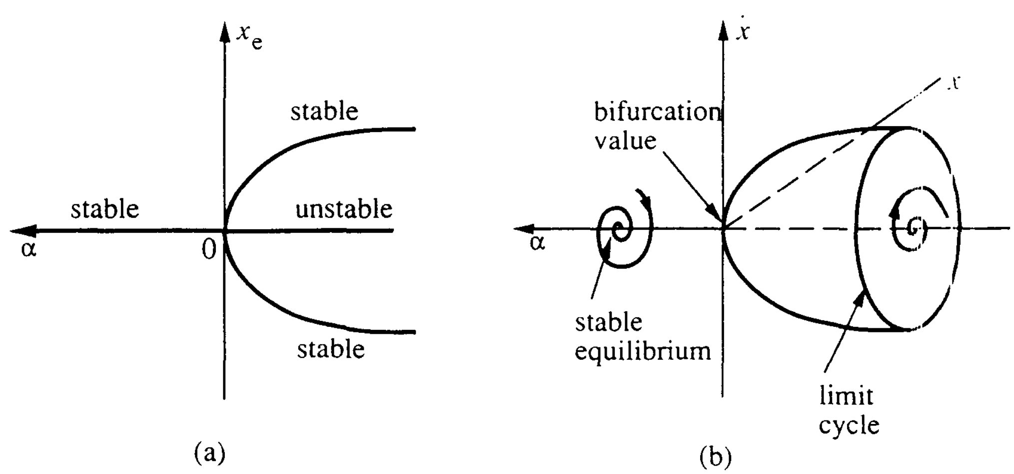

Figure 1.5 : (a) a pitchfork bifurcation; (b) a Hopf bifurcation

Chaos

For stable linear systems, small differences in initial conditions can only cause small differences in output. Nonlinear systems, however, can display a phenomenon called *chaos*, by which we mean that the system output is extremely sensitive to initial conditions. The essential feature of chaos is the unpredictability of the system output. Even if we have an exact model of a nonlinear system and an extremely accurate computer, the system's response in the long-run still cannot be well predicted.

Chaos must be distinguished from random motion. In random motion, the system model or input contain uncertainty and, as a result, the time variation of the output cannot be predicted exactly (only statistical measures are available). In chaotic motion, on the other hand, the involved problem is deterministic, and there is little uncertainty in system model, input, or initial conditions.

As an example of chaotic behavior, let us consider the simple nonlinear system

$$\ddot{x} + 0.1\dot{x} + x^5 = 6 \sin t$$

which may represent a lightly-damped, sinusoidally forced mechanical structure undergoing large elastic deflections. Figure 1.6 shows the responses of the system corresponding to two almost identical initial conditions, namely $x(0) = 2, \dot{x}(0) = 3$ (thick line) and $x(0) = 2.01, \dot{x}(0) = 3.01$ (thin line). Due to the presence of the strong nonlinearity in x^5, the two responses are radically different after some time.

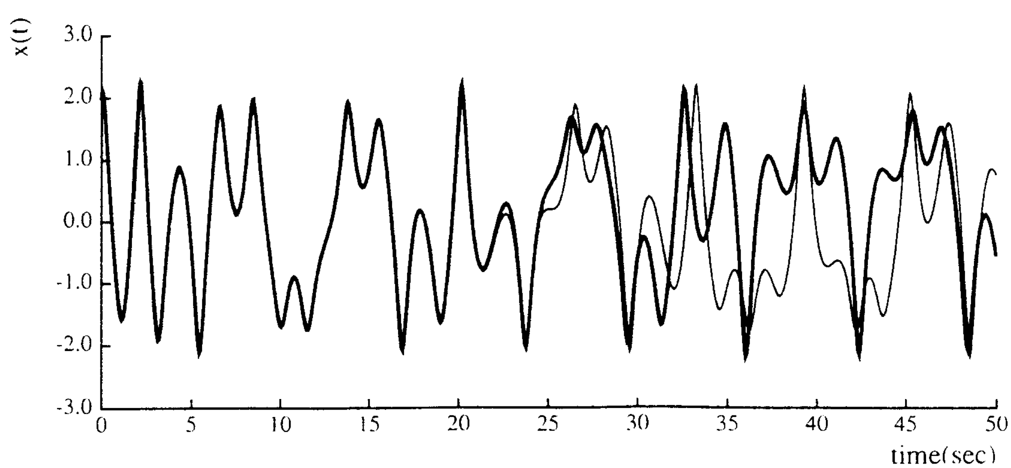

Figure 1.6 : Chaotic behavior of a nonlinear system

Chaotic phenomena can be observed in many physical systems. The most commonly seen physical problem is turbulence in fluid mechanics (such as the swirls of our incense stick). Atmospheric dynamics also display clear chaotic behavior, thus making long-term weather prediction impossible. Some mechanical and electrical systems known to exhibit chaotic vibrations include buckled elastic structures, mechanical systems with play or backlash, systems with aeroelastic dynamics, wheel-rail dynamics in railway systems, and, of course, feedback control devices.

Chaos occurs mostly in *strongly* nonlinear systems. This implies that, for a given system, if the initial condition or the external input cause the system to operate in a highly nonlinear region, it increases the possibility of generating chaos. Chaos cannot occur in linear systems. Corresponding to a sinusoidal input of arbitrary magnitude, the linear system response is always a sinusoid of the same frequency. By contrast, the output of a given nonlinear system may display sinusoidal, periodic, or chaotic behaviors, depending on the initial condition and the input magnitude.

In the context of feedback control, it is of course of interest to know when a nonlinear system will get into a chaotic mode (so as to avoid it) and, in case it does, how to recover from it. Such problems are the object of active research.

Other behaviors

Other interesting types of behavior, such as jump resonance, subharmonic generation, asynchronous quenching, and frequency-amplitude dependence of free vibrations, can also occur and become important in some system studies. However, the above description should provide ample evidence that nonlinear systems can have considerably richer and more complex behavior than linear systems.

1.3 An Overview of the Book

Because nonlinear systems can have much richer and more complex behaviors than linear systems, their analysis is much more difficult. Mathematically, this is reflected in two aspects. First, nonlinear equations, unlike linear ones, cannot in general be solved analytically, and therefore a complete understanding of the behavior of a nonlinear system is very difficult. Second, powerful mathematical tools like Laplace and Fourier transforms do not apply to nonlinear systems.

As a result, there are no systematic tools for predicting the behavior of nonlinear systems, nor are there systematic procedures for designing nonlinear control systems. Instead, there is a rich inventory of powerful analysis and design tools, each best applicable to particular classes of nonlinear control problems. It is the objective of this book to present these various tools, with particular emphasis on their powers and limitations, and on how they can be effectively combined.

This book is divided into two major parts. Part I (chapters 2-5) presents the major *analytical* tools that can be used to study a nonlinear system. Part II (chapters 6-9) discusses the major nonlinear controller *design* techniques. Each part starts with a short introduction providing the background for the main issues and techniques to be discussed.

In chapter 2, we further familiarize ourselves with some basic nonlinear system behaviors, by studying second-order systems using the simple graphical tools provided by so-called phase plane analysis. Chapter 3 introduces the most fundamental analysis tool to be used in this book, namely the concept of a Lyapunov function and its use in nonlinear stability analysis. Chapter 4 studies selected advanced topics in stability analysis. Chapter 5 discusses an approximate nonlinear system analysis method, the describing function method, which aims at extending to nonlinear systems some of the desirable and intuitive properties of linear frequency response analysis.

The basic idea of chapter 6 is to study under what conditions the dynamics of a nonlinear system can be algebraically transformed in that of a linear system, on which linear control design techniques can in turn be applied. Chapters 7 and 8 then study how to reduce or practically eliminate the effects of model uncertainties on the stability and performance of feedback controllers for linear or nonlinear systems, using so-called robust and adaptive approaches. Finally, chapter 9 extensively discusses the use of known physical properties to simplify and enhance the design of controllers for complex multi-input nonlinear systems.

The book concentrates on nonlinear systems represented in continuous-time form. Even though most control systems are implemented digitally, nonlinear physical systems are continuous in nature and are hard to meaningfully discretize, while digital control systems may be treated as continuous-time systems in analysis and design if high sampling rates are used. Given the availability of cheap computation, the most common practical case when it may be advantageous to consider sampling explicitly is when *measurements* are sparse, as *e.g.*, in the case of underwater vehicles using acoustic navigation. Some practical issues involved in the digital implementation of controllers designed from continuous-time formulations are discussed in the introduction to Part II.

1.4 Notes and References

Detailed discussions of bifurcations and chaos can be found, *e.g.*, in [Guckenheimer and Holmes, 1983] and in [Thompson and Stewart, 1986], from which the example of Figure 1.6 is adapted.

Part I
Nonlinear Systems Analysis

The objective of this part is to present various tools available for analyzing nonlinear control systems. The study of these nonlinear analysis techniques is important for a number of reasons. First, theoretical analysis is usually the least expensive way of exploring a system's characteristics. Second, simulation, though very important in nonlinear control, has to be guided by theory. Blind simulation of nonlinear systems is likely to produce few results or misleading results. This is especially true given the great richness of behavior that nonlinear systems can exhibit, depending on initial conditions and inputs. Third, the design of nonlinear controllers is always based on analysis techniques. Since design methods are usually based on analysis methods, it is almost impossible to master the design methods without first studying the analysis tools. Furthermore, analysis tools also allow us to assess control designs after they have been made, and, in case of inadequate performance, they may also suggest directions of modifying the control designs.

It should not come as a surprise that no universal technique has been devised for the analysis of all nonlinear control systems. In linear control, one can analyze a system in the time domain or in the frequency domain. However, for nonlinear control systems, none of these standard approaches can be used, since direct solution of nonlinear differential equations is generally impossible, and frequency domain transformations do not apply.

While the analysis of nonlinear control systems is difficult, serious efforts have been made to develop appropriate theoretical tools for it. Many methods of nonlinear control system analysis have been proposed. Let us briefly describe some of these methods before discussing their details in the following chapters.

Phase plane analysis

Phase plane analysis, discussed in chapter 2, is a graphical method of studying second-order nonlinear systems. Its basic idea is to solve a second order differential equation graphically, instead of seeking an analytical solution. The result is a family of system motion trajectories on a two-dimensional plane, called the phase plane, which allow us to visually observe the motion patterns of the system. While phase plane analysis has a number of important advantages, it has the fundamental disadvantage of being applicable only to systems which can be well approximated by a second-order dynamics. Because of its graphical nature, it is frequently used to provide intuitive insights about nonlinear effects.

Lyapunov theory

Basic Lyapunov theory comprises two methods introduced by Lyapunov, the indirect method and the direct method. The indirect method, or linearization method, states that the stability properties of a nonlinear system in the close vicinity of an equilibrium point are essentially the same as those of its linearized approximation. The method serves as the theoretical justification for using linear control for physical systems, which are always inherently nonlinear. The direct method is a powerful tool for nonlinear system analysis, and therefore the so-called Lyapunov analysis often actually refers to the direct method. The direct method is a generalization of the energy concepts associated with a mechanical system: the motion of a mechanical system is stable if its total mechanical energy decreases all the time. In using the direct method to analyze the stability of a nonlinear system, the idea is to construct a scalar energy-like function (a Lyapunov function) for the system, and to see whether it decreases. The power of this method comes from its generality: it is applicable to all kinds of control systems, be they time-varying or time-invariant, finite dimensional or infinite dimensional. Conversely, the limitation of the method lies in the fact that it is often difficult to find a Lyapunov function for a given system.

Although Lyapunov's direct method is originally a method of stability analysis, it can be used for other problems in nonlinear control. One important application is the design of nonlinear controllers. The idea is to somehow formulate a scalar positive function of the system states, and then choose a control law to make this function decrease. A nonlinear control system thus designed will be guaranteed to be stable. Such a design approach has been used to solve many complex design problems, *e.g.*, in robotics and adaptive control. The direct method can also be used to estimate the performance of a control system and study its robustness. The important subject of Lyapunov analysis is studied in chapters 3 and 4, with chapter 3 presenting the main concepts and results in Lyapunov theory, and chapter 4 discussing some advanced topics.

Describing functions

The describing function method is an approximate technique for studying nonlinear systems. The basic idea of the method is to approximate the nonlinear components in nonlinear control systems by linear "equivalents", and then use frequency domain techniques to analyze the resulting systems. Unlike the phase plane method, it is not restricted to second-order systems. Unlike Lyapunov methods, whose applicability to a specific system hinges on the success of a trial-and-error search for a Lyapunov function, its application is straightforward for nonlinear systems satisfying some easy-to-check conditions.

The method is mainly used to predict limit cycles in nonlinear systems. Other applications include the prediction of subharmonic generation and the determination of system response to sinusoidal excitation. The method has a number of advantages. First, it can deal with low order and high order systems with the same straightforward procedure. Second, because of its similarity to frequency-domain analysis of linear systems, it is conceptually simple and physically appealing, allowing users to exercise their physical and engineering insights about the control system. Third, it can deal with the "hard nonlinearities" frequently found in control systems without any difficulty. As a result, it is an important tool for practical problems of nonlinear control analysis and design. The disadvantages of the method are linked to its approximate nature, and include the possibility of inaccurate predictions (false predictions may be made if certain conditions are not satisfied) and restrictions on the systems to which it applies (for example, it has difficulties in dealing with systems with multiple nonlinearities).

Chapter 2
Phase Plane Analysis

Phase plane analysis is a graphical method for studying second-order systems, which was introduced well before the turn of the century by mathematicians such as Henri Poincare. The basic idea of the method is to generate, in the state space of a second-order dynamic system (a two-dimensional plane called the phase plane), motion trajectories corresponding to various initial conditions, and then to examine the qualitative features of the trajectories. In such a way, information concerning stability and other motion patterns of the system can be obtained. In this chapter, our objective is to gain familiarity with nonlinear systems through this simple graphical method.

Phase plane analysis has a number of useful properties. First, as a graphical method, it allows us to visualize what goes on in a nonlinear system starting from various initial conditions, without having to solve the nonlinear equations analytically. Second, it is not restricted to small or smooth nonlinearities, but applies equally well to strong nonlinearities and to "hard" nonlinearities. Finally, some practical control systems can indeed be adequately approximated as second-order systems, and the phase plane method can be used easily for their analysis. Conversely, of course, the fundamental disadvantage of the method is that it is restricted to second-order (or first-order) systems, because the graphical study of higher-order systems is computationally and geometrically complex.

2.1 Concepts of Phase Plane Analysis

2.1.1 Phase Portraits

The phase plane method is concerned with the graphical study of second-order autonomous systems described by

$$\dot{x}_1 = f_1(x_1, x_2) \tag{2.1a}$$

$$\dot{x}_2 = f_2(x_1, x_2) \tag{2.1b}$$

where x_1 and x_2 are the states of the system, and f_1 and f_2 are nonlinear functions of the states. Geometrically, the state space of this system is a plane having x_1 and x_2 as coordinates. We will call this plane the *phase plane*.

Given a set of initial conditions $\mathbf{x}(0) = \mathbf{x}_o$, Equation (2.1) defines a solution $\mathbf{x}(t)$. With time t varied from zero to infinity, the solution $\mathbf{x}(t)$ can be represented geometrically as a curve in the phase plane. Such a curve is called a phase plane *trajectory*. A family of phase plane trajectories corresponding to various initial conditions is called a *phase portrait* of a system.

To illustrate the concept of phase portrait, let us consider the following simple system.

Example 2.1: Phase portrait of a mass-spring system

The governing equation of the mass-spring system in Figure 2.1(a) is the familiar linear second-order differential equation

$$\ddot{x} + x = 0 \tag{2.2}$$

Assume that the mass is initially at rest, at length x_o. Then the solution of the equation is

$$x(t) = x_o \cos t$$

$$\dot{x}(t) = -\, x_o \sin t$$

Eliminating time t from the above equations, we obtain the equation of the trajectories

$$x^2 + \dot{x}^2 = x_o^{\,2}$$

This represents a circle in the phase plane. Corresponding to different initial conditions, circles of different radii can be obtained. Plotting these circles on the phase plane, we obtain a phase portrait for the mass-spring system (Figure 2.1.b). □

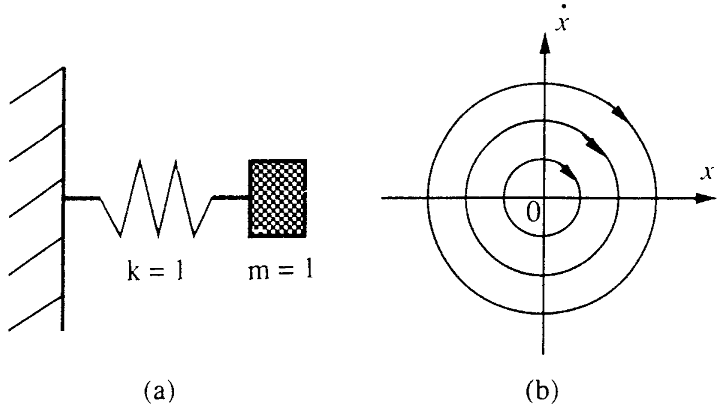

Figure 2.1 : A mass-spring system and its phase portrait

The power of the phase portrait lies in the fact that once the phase portrait of a system is obtained, the nature of the system response corresponding to various initial conditions is directly displayed on the phase plane. In the above example, we easily see that the system trajectories neither converge to the origin nor diverge to infinity. They simply circle around the origin, indicating the marginal nature of the system's stability.

A major class of second-order systems can be described by differential equations of the form

$$\ddot{x} + f(x, \dot{x}) = 0 \tag{2.3}$$

In state space form, this dynamics can be represented as

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = -f(x_1, x_2)$$

with $x_1 = x$ and $x_2 = \dot{x}$. Most second-order systems in practice, such as mass-damper-spring systems in mechanics, or resistor-coil-capacitor systems in electrical engineering, can be represented in or transformed into this form. For these systems, the states are x and its derivative $\dot{x}$. Traditionally, the phase plane method is developed for the dynamics (2.3), and the phase plane is defined as the plane having x and $\dot{x}$ as coordinates. But it causes no difficulty to extend the method to more general dynamics of the form (2.1), with the (x_1 , x_2) plane as the phase plane, as we do in this chapter.

2.1.2 Singular Points

An important concept in phase plane analysis is that of a singular point. A singular point is an equilibrium point in the phase plane. Since an equilibrium point is defined as a point where the system states can stay forever, this implies that $\dot{\mathbf{x}} = \mathbf{0}$, and using (2.1),

$$f_1(x_1, x_2) = 0 \qquad f_2(x_1, x_2) = 0 \tag{2.4}$$

The values of the equilibrium states can be solved from (2.4).

For a linear system, there is usually only one singular point (although in some cases there can be a *continuous* set of singular points, as in the system $\ddot{x} + \dot{x} = 0$, for which all points on the real axis are singular points). However, a nonlinear system often has more than one isolated singular point, as the following example shows.

Example 2.2: A nonlinear second-order system

Consider the system

$$\ddot{x} + 0.6\,\dot{x} + 3\,x + x^2 = 0$$

whose phase portrait is plotted in Figure 2.2. The system has two singular points, one at (0, 0) and the other at (−3, 0). The motion patterns of the system trajectories in the vicinity of the two singular points have different natures. The trajectories move towards the point $x = 0$ while moving away from the point $x = -3$. □

One may wonder why an equilibrium point of a second-order system is called a *singular* point. To answer this, let us examine the slope of the phase trajectories. From (2.1), the slope of the phase trajectory passing through a point (x_1, x_2) is determined by

$$\frac{dx_2}{dx_1} = \frac{f_2(x_1, x_2)}{f_1(x_1, x_2)} \tag{2.5}$$

With the functions f_1 and f_2 assumed to be single valued, there is usually a definite value for this slope at any given point in phase plane. This implies that the phase trajectories will not intersect. At singular points, however, the value of the slope is 0/0, *i.e.*, the slope is indeterminate. Many trajectories may intersect at such points, as seen from Figure 2.2. This indeterminacy of the slope accounts for the adjective "singular".

Singular points are very important features in the phase plane. Examination of the singular points can reveal a great deal of information about the properties of a

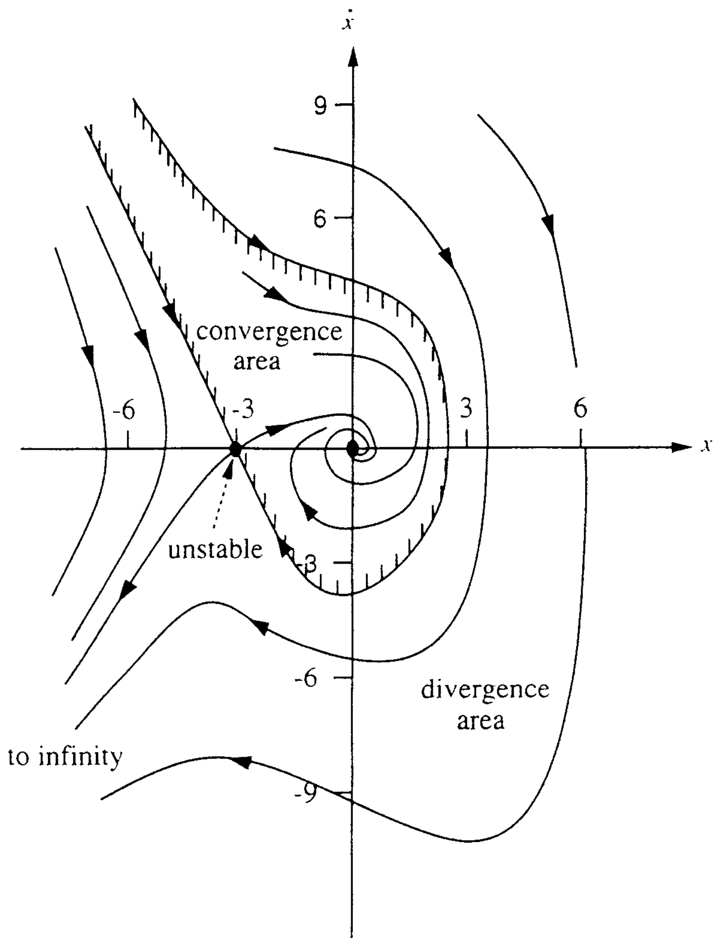

Figure 2.2 : The phase portrait of a nonlinear system

system. In fact, the stability of linear systems is uniquely characterized by the nature of their singular points. For nonlinear systems, besides singular points, there may be more complex features, such as limit cycles. These issues will be discussed in detail in sections 2.3 and 2.4.

Note that, although the phase plane method is developed primarily for second-order systems, it can also be applied to the analysis of first-order systems of the form

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

The idea is still to plot $\dot{x}$ with respect to x in the phase plane. The difference now is that the phase portrait is composed of a single trajectory.

Example 2.3: A first-order system

Consider the system

$$\dot{x} = -4x + x^3$$

There are three singular points, defined by $-4x + x^3 = 0$, namely, $x = 0, -2,$ and 2. The phase-portrait of the system consists of a single trajectory, and is shown in Figure 2.3. The arrows in the figure denote the direction of motion, and whether they point toward the left or the right at a particular point is determined by the sign of $\dot{x}$ at that point. It is seen from the phase portrait of this system that the equilibrium point $x = 0$ is stable, while the other two are unstable. □

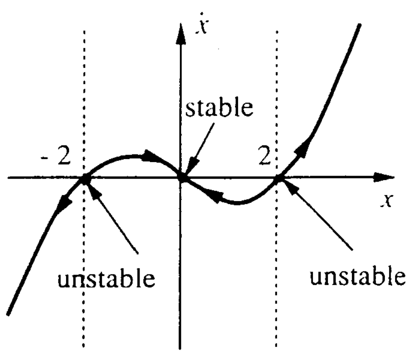

Figure 2.3 : Phase trajectory of a first-order system

2.1.3 Symmetry in Phase Plane Portraits

A phase portrait may have *a priori* known symmetry properties, which can simplify its generation and study. If a phase portrait is symmetric with respect to the x_1 or the x_2 axis, one only needs in practice to study half of it. If a phase portrait is symmetric with respect to both the x_1 and x_2 axes, only one quarter of it has to be explicitly considered.

Before generating a phase portrait itself, we can determine its symmetry properties by examining the system equations. Let us consider the second-order dynamics (2.3). The slope of trajectories in the phase plane is of the form

$$\frac{dx_2}{dx_1} = -\frac{f(x_1, x_2)}{\dot{x}}$$

Since symmetry of the phase portraits also implies symmetry of the slopes (equal in absolute value but opposite in sign), we can identify the following situations

Symmetry about the x_1 axis: The condition is

$$f(x_1, x_2) = f(x_1, -x_2)$$

This implies that the function f should be even in x_2. The mass-spring system in Example 2.1 satisfies this condition. Its phase portrait is seen to be symmetric about the x_1 axis.

Symmetry about the x_2 axis: Similarly,

$$f(x_1, x_2) = -f(-x_1, x_2)$$

implies symmetry with respect to the x_2 axis. The mass-spring system also satisfies this condition.

Symmetry about the origin: When

$$f(x_1, x_2) = -f(-x_1, -x_2)$$

the phase portrait of the system is symmetric about the origin.

2.2 Constructing Phase Portraits

Today, phase portraits are routinely computer-generated. In fact, it is largely the advent of the computer in the early 1960's, and the associated ease of quickly generating phase portraits, which spurred many advances in the study of complex nonlinear dynamic behaviors such as chaos. However, of course (as *e.g.*, in the case of root locus for linear systems), it is still practically useful to learn how to roughly sketch phase portraits or quickly verify the plausibility of computer outputs.

There are a number of methods for constructing phase plane trajectories for linear or nonlinear systems, such as the so-called analytical method, the method of isoclines, the delta method, Lienard's method, and Pell's method. We shall discuss two of them in this section, namely, the analytical method and the method of isoclines. These methods are chosen primarily because of their relative simplicity. The analytical method involves the analytical solution of the differential equations describing the systems. It is useful for some special nonlinear systems, particularly piece-wise linear systems, whose phase portraits can be constructed by piecing together the phase portraits of the related linear systems. The method of isoclines is a graphical method which can conveniently be applied to construct phase portraits for systems which cannot be solved analytically, which represent by far the most common case.

ANALYTICAL METHOD

There are two techniques for generating phase plane portraits analytically. Both techniques lead to a functional relation between the two phase variables x and x_2 in the form

$$g(x_1, x_2, c) = 0 \tag{2.6}$$

where the constant c represents the effects of initial conditions (and, possibly, of external input signals). Plotting this relation in the phase plane for different initial conditions yields a phase portrait.

The first technique involves solving equations (2.1) for x_1 and x_2 as functions of time t, *i.e.*,

$$x_1(t) = g_1(t) \qquad x_2(t) = g_2(t)$$

and then eliminating time t from these equations, leading to a functional relation in the form of (2.6). This technique was already illustrated in Example 2.1.

The second technique, on the other hand, involves directly eliminating the time variable, by noting that

$$\frac{dx_2}{dx_1} = \frac{f_2(x_1, x_2)}{f_1(x_1, x_2)}$$

and then solving this equation for a functional relation between x_1 and x_2. Let us use this technique to solve the mass-spring equation again.

Example 2.4: Mass-spring system

By noting that $\ddot{x} = (d\dot{x}/dx)(dx/dt)$, we can rewrite (2.2) as

$$\dot{x}\frac{d\dot{x}}{dx} + x = 0$$

Integration of this equation yields

$$\dot{x}^2 + x^2 = x_o^2$$

□

One sees that the second technique is more straightforward in generating the equations for the phase plane trajectories.

Most nonlinear systems cannot be easily solved by either of the above two techniques. However, for piece-wise linear systems, an important class of nonlinear systems, this method can be conveniently used, as the following example shows.

Example 2.5: A satellite control system

Figure 2.4 shows the control system for a simple satellite model. The satellite, depicted in Figure 2.5(a), is simply a rotational unit inertia controlled by a pair of thrusters, which can provide either a positive constant torque U (positive firing) or a negative torque $-U$ (negative firing). The purpose of the control system is to maintain the satellite antenna at a zero angle by appropriately firing the thrusters. The mathematical model of the satellite is

$$\ddot{\theta} = u$$

where u is the torque provided by the thrusters and θ is the satellite angle.

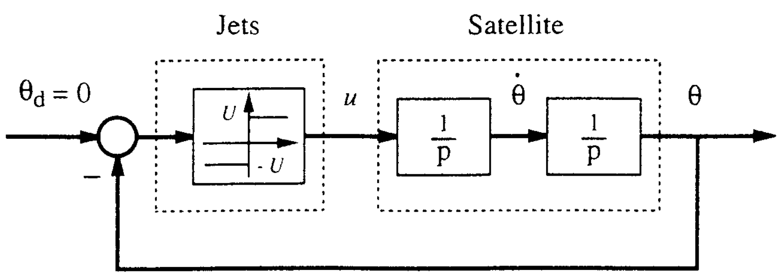

Figure 2.4 : Satellite control system

Let us examine on the phase plane the behavior of the control system when the thrusters are fired according to the control law

$$u(t) = \begin{cases} -U & \text{if } \theta > 0 \\ U & \text{if } \theta < 0 \end{cases} \tag{2.7}$$

which means that the thrusters push in the counterclockwise direction if θ is positive, and vice versa.

As the first step of the phase portrait generation, let us consider the phase portrait when the thrusters provide a positive torque U. The dynamics of the system is

$$\ddot{\theta} = U$$

which implies that $\dot{\theta}\, d\dot{\theta} = U\, d\theta$. Therefore, the phase trajectories are a family of parabolas defined by

$$\dot{\theta}^2 = 2U\theta + c_1$$

where c_1 is a constant. The corresponding phase portrait of the system is shown in Figure 2.5(b).

When the thrusters provide a negative torque $-U$, the phase trajectories are similarly found to be

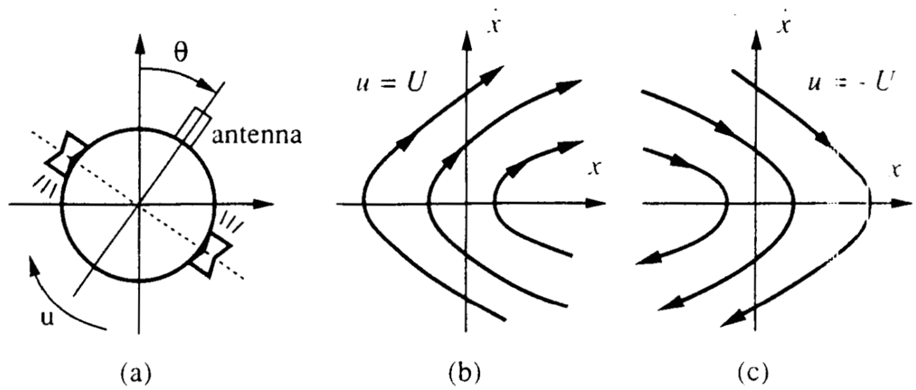

Figure 2.5 : Satellite control using on-off thrusters

$$\dot{\theta}^2 = -2Ux + c_1$$

with the corresponding phase portrait shown in Figure 2.5(c).

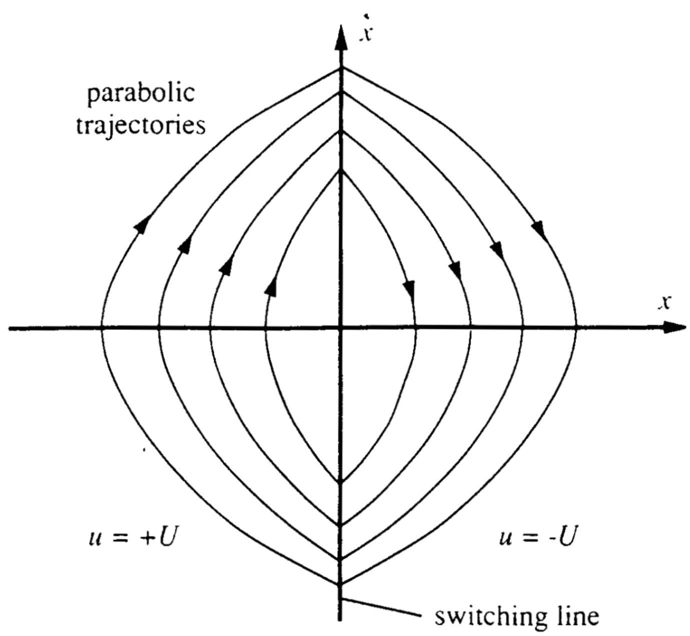

Figure 2.6 : Complete phase portrait of the control system

The complete phase portrait of the closed-loop control system can be obtained simply by connecting the trajectories on the left half of the phase plane in 2.5(b) with those on the right half of the phase plane in 2.5(c), as shown in Figure 2.6. The vertical axis represents a switching line, because the control input and thus the phase trajectories are switched on that line. It is interesting to see that, starting from a nonzero initial angle, the satellite will oscillate in periodic motions under the action of the jets. One concludes from this phase portrait that the system is marginally stable, similarly to the mass-spring system in Example 2.1. Convergence of the system to the zero angle can be obtained by adding rate feedback (Exercise 2.4). □

THE METHOD OF ISOCLINES

The basic idea in this method is that of isoclines. Consider the dynamics in (2.1). At a point (x_1, x_2) in the phase plane, the slope of the tangent to the trajectory can be determined by (2.5). An isocline is defined to be the locus of the points with a given tangent slope. An isocline with slope α is thus defined to be

$$\frac{dx_2}{dx_1} = \frac{f_2(x_1, x_2)}{f_1(x_1, x_2)} = \alpha$$

This is to say that points on the curve

$$f_2(x_1, x_2) = \alpha f_1(x_1, x_2)$$

all have the same tangent slope α.

In the method of isoclines, the phase portrait of a system is generated in two steps. In the first step, a field of directions of tangents to the trajectories is obtained. In the second step, phase plane trajectories are formed from the field of directions .

Let us explain the isocline method on the mass-spring system in (2.2). The slope of the trajectories is easily seen to be

$$\frac{dx_2}{dx_1} = -\frac{x_1}{x_2}$$

Therefore, the isocline equation for a slope α is

$$x_1 + \alpha\, x_2 = 0$$

i.e., a straight line. Along the line, we can draw a lot of short line segments with slope α. By taking α to be different values, a set of isoclines can be drawn, and a field of directions of tangents to trajectories are generated, as shown in Figure 2.7. To obtain trajectories from the field of directions, we assume that the the tangent slopes are locally constant. Therefore, a trajectory starting from any point in the plane can be found by connecting a sequence of line segments.

Let us use the method of isoclines to study the Van der Pol equation, a nonlinear equation.

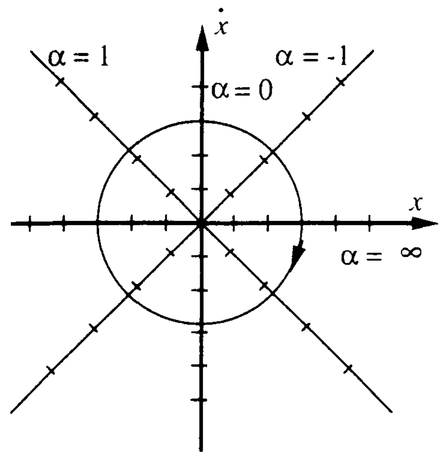

Figure 2.7 : Isoclines for the mass-spring system

Example 2.6: The Van der Pol equation

For the Van der Pol equation

$$\ddot{x} + 0.2\,(x^2 - 1)\,\dot{x} + x = 0$$

an isocline of slope α is defined by

$$\frac{d\dot{x}}{dx} = -\frac{0.2\,(x^2 - 1)\,\dot{x} + x}{\dot{x}} = \alpha$$

Therefore, the points on the curve

$$0.2\,(x^2 - 1)\,\dot{x} + x + \alpha\,\dot{x} = 0$$

all have the same slope α.

By taking α of different values, different isoclines can be obtained, as plotted in Figure 2.8. Short line segments are drawn on the isoclines to generate a field of tangent directions. The phase portraits can then be obtained, as shown in the plot. It is interesting to note that there exists a closed curve in the portrait, and the trajectories starting from both outside and inside converge to this curve. This closed curve corresponds to a limit cycle, as will be discussed further in section 2.5. □

Note that the same scales should be used for the x_1 axis and x_2 axis of the phase plane, so that the derivative dx_2/dx_1 equals the geometric slope of the trajectories. Also note that, since in the second step of phase portrait construction we essentially assume that the slope of the phase plane trajectories is locally constant, more isoclines should be plotted in regions where the slope varies quickly, to improve accuracy.

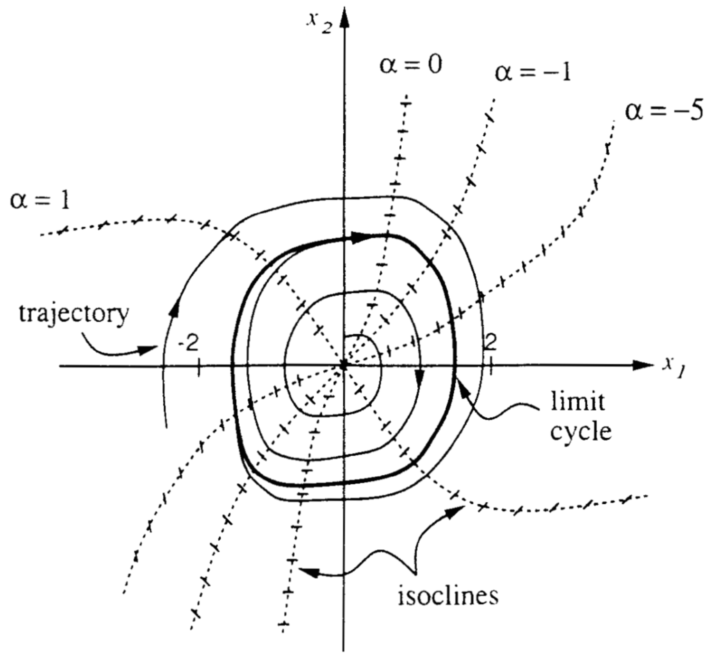

Figure 2.8 : Phase portrait of the Van der Pol equation

2.3 Determining Time from Phase Portraits

Note that time t does not explicitly appear in the phase plane having x_1 and x_2 as coordinates. However, in some cases, we might be interested in the time information. For example, one might want to know the time history of the system states starting from a specific initial point. Another relevant situation is when one wants to know how long it takes for the system to move from a point to another point in a phase plane trajectory. We now describe two techniques for computing time history from phase portraits. Both techniques involve a step-by step procedure for recovering time.

Obtaining time from $\Delta t \approx \Delta x / \dot{x}$

In a short time Δt, the change of x is approximately

$$\Delta x \approx \dot{x} \Delta t \tag{2.8}$$

where $\dot{x}$ is the velocity corresponding to the increment Δx. Note that for a Δx of finite magnitude, the average value of velocity during a time increment should be used to improve accuracy. From (2.8), the length of time corresponding to the increment Δx

is

$$\Delta t \approx \frac{\Delta x}{\dot{x}}$$

The above reasoning implies that, in order to obtain the time corresponding to the motion from one point to another point along a trajectory, one should divide the corresponding part of the trajectory into a number of small segments (not necessarily equally spaced), find the time associated with each segment, and then add up the results. To obtain the time history of states corresponding to a certain initial condition, one simply computes the time t for each point on the phase trajectory, and then plots x with respect to t and $\dot{x}$ with respect to t.

Obtaining time from $t = \int (1/\dot{x})\, dx$

Since $\dot{x} = dx/dt$, we can write $dt = dx/\dot{x}$. Therefore,

$$t - t_o = \int_{x_o}^{x} (1/\dot{x})\, dx$$

where x corresponds to time t and x_o corresponds to time t_o. This equation implies that, if we plot a phase plane portrait with new coordinates x and $(1/\dot{x})$, then the area under the resulting curve is the corresponding time interval.

2.4 Phase Plane Analysis of Linear Systems

In this section, we describe the phase plane analysis of linear systems. Besides allowing us to visually observe the motion patterns of linear systems, this will also help the development of nonlinear system analysis in the next section, because a nonlinear systems behaves similarly to a linear system around each equilibrium point.

The general form of a linear second-order system is

$$\dot{x}_1 = a x_1 + b x_2 \tag{2.9a}$$

$$\dot{x}_2 = c x_1 + d x_2 \tag{2.9b}$$

To facilitate later discussions, let us transform this equation into a scalar second-order differential equation. Note from (2.9a) and (2.9b) that

$$b\,\dot{x}_2 = b\,c x_1 + d(\dot{x}_1 - a x_1)$$

Consequently, differentiation of (2.9a) and then substitution of (2.9b) leads to

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

Therefore, we will simply consider the second-order linear system described by

$$\ddot{x} + a\dot{x} + bx = 0 \tag{2.10}$$

To obtain the phase portrait of this linear system, we first solve for the time history

$$x(t) = k_1 e^{\lambda_1 t} + k_2 e^{\lambda_2 t} \qquad \text{for } \lambda_1 \neq \lambda_2 \tag{2.11a}$$

$$x(t) = k_1 e^{\lambda_1 t} + k_2 t\, e^{\lambda_1 t} \qquad \text{for } \lambda_1 = \lambda_2 \tag{2.11b}$$

where the constants λ_1 and λ_2 are the solutions of the characteristic equation

$$s^2 + as + b = (s - \lambda_1)(s - \lambda_2) = 0$$

The roots λ_1 and λ_2 can be explicitly represented as

$$\lambda_1 = (-a + \sqrt{a^2 - 4b}\,)/2 \qquad \lambda_2 = (-a - \sqrt{a^2 - 4b}\,)/2$$

For linear systems described by (2.10), there is only one singular point (assuming $b \neq 0$), namely the origin. However, the trajectories in the vicinity of this singularity point can display quite different characteristics, depending on the values of a and b. The following cases can occur

1. λ_1 and λ_2 are both real and have the same sign (positive or negative)
2. λ_1 and λ_2 are both real and have opposite signs
3. λ_1 and λ_2 are complex conjugate with non-zero real parts
4. λ_1 and λ_2 are complex conjugates with real parts equal to zero

We now briefly discuss each of the above four cases.

STABLE OR UNSTABLE NODE

The first case corresponds to a *node*. A node can be stable or unstable. If the eigenvalues are negative, the singularity point is called a *stable node* because both $x(t)$ and $\dot{x}(t)$ converge to zero exponentially, as shown in Figure 2.9(a). If both eigenvalues are positive, the point is called an *unstable node*, because both $x(t)$ and $\dot{x}(t)$ diverge from zero exponentially, as shown in Figure 2.9(b). Since the eigenvalues are real, there is no oscillation in the trajectories.

SADDLE POINT

The second case (say $\lambda_1 < 0$ and $\lambda_2 > 0$) corresponds to a *saddle point* (Figure 2.9(c)). The phase portrait of the system has the interesting "saddle" shape shown in Figure 2.9(c). Because of the unstable pole λ_2, almost all of the system trajectories diverge to infinity. In this figure, one also observes two straight lines passing through the origin. The diverging line (with arrows pointing to infinity) corresponds to initial conditions which make k_2 (*i.e.*, the unstable component) equal zero. The converging straight line corresponds to initial conditions which make k_1 equal zero.

STABLE OR UNSTABLE FOCUS

The third case corresponds to a focus. A *stable focus* occurs when the real part of the eigenvalues is negative, which implies that $x(t)$ and $\dot{x}(t)$ both converge to zero. The system trajectories in the vicinity of a stable focus are depicted in Figure 2.9(d). Note that the trajectories encircle the origin one or more times before converging to it, unlike the situation for a stable node. If the real part of the eigenvalues is positive, then $x(t)$ and $\dot{x}(t)$ both diverge to infinity, and the singularity point is called an *unstable focus*. The trajectories corresponding to an unstable focus are sketched in Figure 2.9(e).

CENTER POINT

The last case corresponds to a center point, as shown in Figure 2.9(f). The name comes from the fact that all trajectories are ellipses and the singularity point is the center of these ellipses. The phase portrait of the undamped mass-spring system belongs to this category.

Note that the stability characteristics of linear systems are uniquely determined by the nature of their singularity points. This, however, is not true for nonlinear systems.

2.5 Phase Plane Analysis of Nonlinear Systems

In discussing the phase plane analysis of nonlinear systems, two points should be kept in mind. Phase plane analysis of nonlinear systems is related to that of linear systems, because the local behavior of a nonlinear system can be approximated by the behavior of a linear system. Yet, nonlinear systems can display much more complicated patterns in the phase plane, such as multiple equilibrium points and limit cycles. We now discuss these points in more detail.

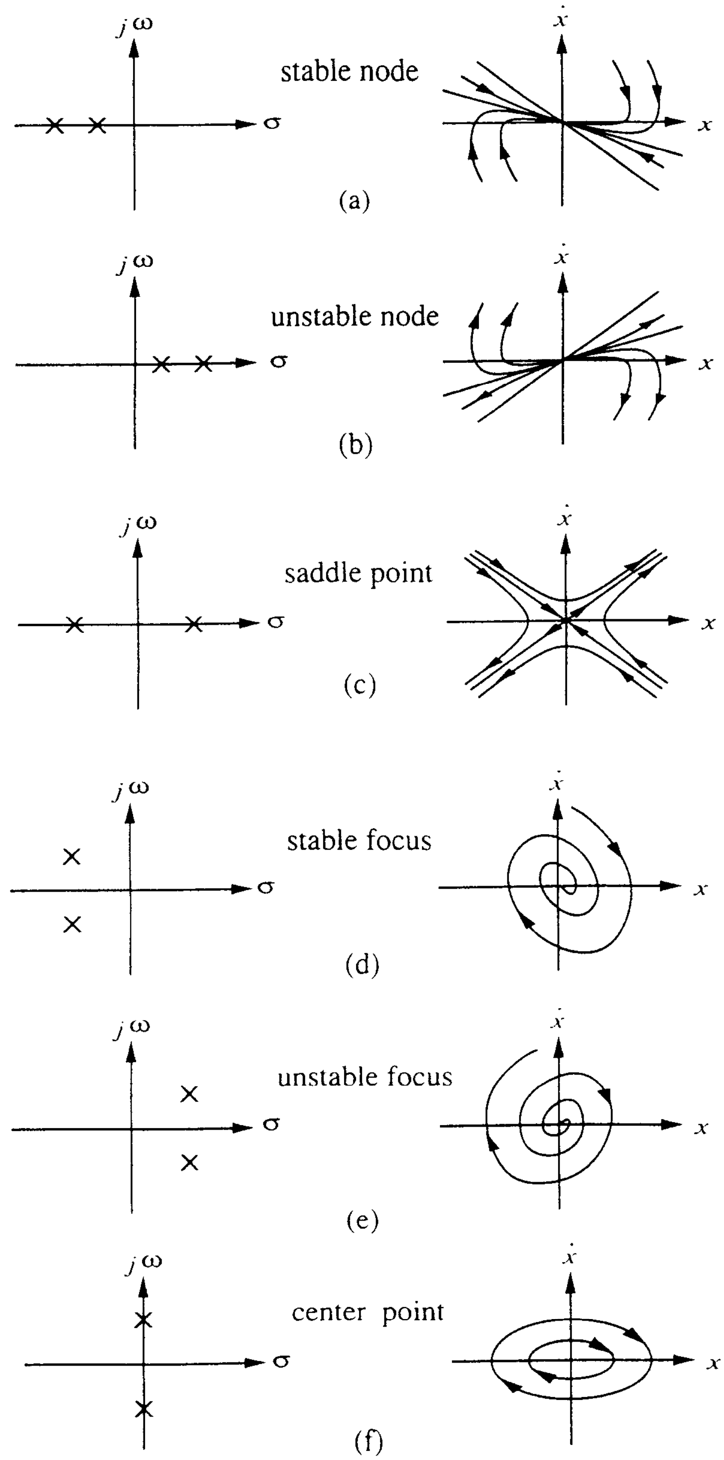

Figure 2.9 : Phase-portraits of linear systems

LOCAL BEHAVIOR OF NONLINEAR SYSTEMS

In the phase portrait of Figure 2.2, one notes that, in contrast to linear systems, there are two singular points, (0, 0) and (−3, 0). However, we also note that the features of the phase trajectories in the neighborhood of the two singular points look very much like those of linear systems, with the first point corresponding to a stable focus and the second to a saddle point. This similarity to a linear system in the local region of each singular point can be formalized by linearizing the nonlinear system, as we now discuss.

If the singular point of interest is not at the origin, by defining the difference between the original state and the singular point as a new set of state variables, one can always shift the singular point to the origin. Therefore, without loss of generality, we may simply consider Equation (2.1) with a singular point at 0. Using Taylor expansion, Equations (2.1a) and (2.1b) can be rewritten as

$$\dot{x}_1 = a x_1 + b x_2 + g_1(x_1, x_2)$$

$$\dot{x}_2 = c x_1 + d x_2 + g_2(x_1, x_2)$$

where g_1 and g_2 contain higher order terms.

In the vicinity of the origin, the higher order terms can be neglected, and therefore, the nonlinear system trajectories essentially satisfy the linearized equation

$$\dot{x}_1 = a\, x_1 + b\, x_2$$

$$\dot{x}_2 = c\, x_1 + d\, x_2$$

As a result, the local behavior of the nonlinear system can be approximated by the patterns shown in Figure 2.9.

LIMIT CYCLES

In the phase portrait of the nonlinear Van der Pol equation, shown in Figure 2.8, one observes that the system has an unstable node at the origin. Furthermore, there is a closed curve in the phase portrait. Trajectories inside the curve and those outside the curve all tend to this curve, while a motion started on this curve will stay on it forever, circling periodically around the origin. This curve is an instance of the so-called "limit cycle" phenomenon. Limit cycles are unique features of nonlinear systems.

In the phase plane, *a limit cycle is defined as an isolated closed curve*. The trajectory has to be both closed, indicating the periodic nature of the motion, and isolated, indicating the limiting nature of the cycle (with nearby trajectories converging or diverging from it). Thus, while there are many closed curves in the phase portraits of the mass-spring-damper system in Example 2.1 or the satellite system in Example 2.5, these are not considered limit cycles in this definition, because they are not isolated.

Depending on the motion patterns of the trajectories in the vicinity of the limit cycle, one can distinguish three kinds of limit cycles

1. **Stable Limit Cycles**: all trajectories in the vicinity of the limit cycle converge to it as $t \to \infty$ (Figure 2.10(a));
2. **Unstable Limit Cycles**: all trajectories in the vicinity of the limit cycle diverge from it as $t \to \infty$ (Figure 2.10(b));
3. **Semi-Stable Limit Cycles**: some of the trajectories in the vicinity converge to it, while the others diverge from it as $t \to \infty$ (Figure 2.10(c));

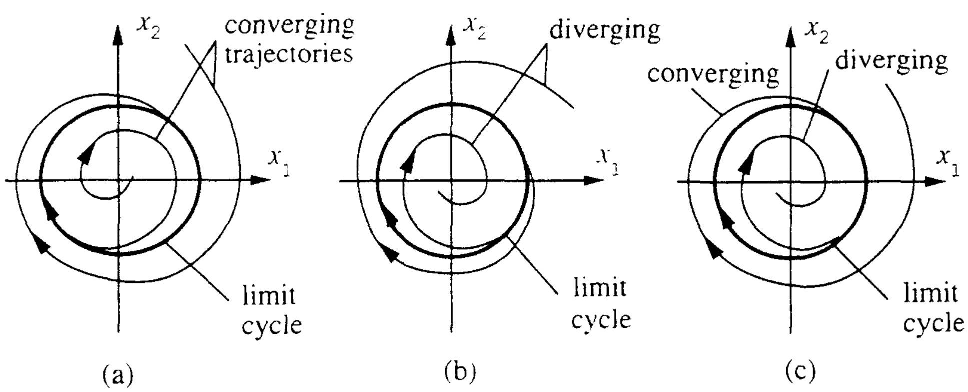

Figure 2.10 : Stable, unstable, and semi-stable limit cycles

As seen from the phase portrait of Figure 2.8, the limit cycle of the Van der Pol equation is clearly stable. Let us consider some additional examples of stable, unstable, and semi-stable limit cycles.

Example 2.7: stable, unstable, and semi-stable limit cycles

Consider the following nonlinear systems

$$\text{(a)} \qquad \dot{x}_1 = x_2 - x_1(x_1^2 + x_2^2 - 1) \qquad \dot{x}_2 = -x_1 - x_2(x_1^2 + x_2^2 - 1) \qquad (2.12)$$

$$\text{(b)} \qquad \dot{x}_1 = x_2 + x_1(x_1^2 + x_2^2 - 1) \qquad \dot{x}_2 = -x_1 + x_2(x_1^2 + x_2^2 - 1) \qquad (2.13)$$

$$\text{(c)} \qquad \dot{x}_1 = x_2 - x_1(x_1^2 + x_2^2 - 1)^2 \qquad \dot{x}_2 = -x_1 - x_2(x_1^2 + x_2^2 - 1)^2 \qquad (2.14)$$

Let us study system (a) first. By introducing polar coordinates

$$r = (x_1^2 + x_2^2)^{1/2} \qquad \theta = \tan^{-1}(x_2/x_1)$$

the dynamic equations (2.12) are transformed as

$$\frac{dr}{dt} = -r(r^2 - 1) \qquad \frac{d\theta}{dt} = -1$$

When the state starts on the unit circle, the above equation shows that $\dot{r}(t) = 0$. Therefore, the state will circle around the origin with a period $1/2\pi$. When $r < 1$, then $\dot{r} > 0$. This implies that the state tends to the circle from inside. When $r > 1$, then $\dot{r} < 0$. This implies that the state tends toward the unit circle from outside. Therefore, the unit circle is a stable limit cycle. This can also be concluded by examining the analytical solution of (2.12)

$$r(t) = \frac{1}{(1 + c_o e^{-2t})^{1/2}} \qquad \theta(t) = \theta_o - t$$

where

$$c_o = \frac{1}{r_o^2} - 1$$

Similarly, one can find that the system (b) has an unstable limit cycle and system (c) has a semi-stable limit cycle. □

2.6 Existence of Limit Cycles

As mentioned in chapter 1, it is of great importance for control engineers to predict the existence of limit cycles in control systems. In this section, we state three simple classical theorems to that effect. These theorems are easy to understand and apply.

The first theorem to be presented reveals a simple relationship between the existence of a limit cycle and the number of singular points it encloses. In the statement of the theorem, we use N to represent the number of nodes, centers, and foci enclosed by a limit cycle, and S to represent the number of enclosed saddle points.

Theorem 2.1 (Poincare) *If a limit cycle exists in the second-order autonomous system (2.1), then* $N = S + 1$.

This theorem is sometimes called the *index theorem*. Its proof is mathematically involved (actually, a family of such proofs led to the development of algebraic topology) and shall be omitted here. One simple inference from this theorem is that a limit cycle must enclose at least one equilibrium point. The theorem's result can be verified easily on Figures 2.8 and 2.10.

The second theorem is concerned with the asymptotic properties of the trajectories of second-order systems.

Theorem 2.2 (Poincare-Bendixson) *If a trajectory of the second-order autonomous system remains in a finite region* Ω, *then one of the following is true:*

(a) the trajectory goes to an equilibrium point

(b) the trajectory tends to an asymptotically stable limit cycle

(c) the trajectory is itself a limit cycle

While the proof of this theorem is also omitted here, its intuitive basis is easy to see, and can be verified on the previous phase portraits.

The third theorem provides a sufficient condition for the non-existence of limit cycles.

Theorem 2.3 (Bendixson) *For the nonlinear system (2.1), no limit cycle can exist in a region* Ω *of the phase plane in which* $\partial f_1/\partial x_1 + \partial f_2/\partial x_2$ *does not vanish and does not change sign.*

Proof: Let us prove this theorem by contradiction. First note that, from (2.5), the equation

$$f_2\, dx_1 - f_1\, dx_2 = 0 \tag{2.15}$$

is satisfied for any system trajectories, including a limit cycle. Thus, along the closed curve L of a limit cycle, we have

$$\int_L (f_1\, dx_2 - f_2\, dx_1) = 0 \tag{2.16}$$

Using Stokes' Theorem in calculus, we have

$$\int_L (f_1\, dx_2 - f_2\, dx_1) = \int\int \left(\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} \right) dx_1\, dx_2 = 0$$

where the integration on the right-hand side is carried out on the area enclosed by the limit cycle.

By Equation (2.16), the left-hand side must equal zero. This, however, contradicts the fact that the right-hand side cannot equal zero because by hypothesis $\partial f_1/\partial x_1 + \partial f_2/\partial x_2$ does not vanish and does not change sign. □

Let us illustrate the result on an example.

Example 2.8: Consider the nonlinear system

$$\dot{x}_1 = g(x_2) + 4x_1x_2^2$$

$$\dot{x}_2 = h(x_1) + 4x_1^2x_2$$

Since

$$\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} = 4(x_1^2 + x_2^2)$$

which is always strictly positive (except at the origin), the system does not have any limit cycles anywhere in the phase plane. □

The above three theorems represent very powerful results. It is important to notice, however, that they have no equivalent in higher-order systems, where exotic asymptotic behaviors other than equilibrium points and limit cycles can occur.

2.7 Summary

Phase plane analysis is a graphical method used to study second-order dynamic systems. The major advantage of the method is that it allows visual examination of the global behavior of systems. The major disadvantage is that it is mainly limited to second-order systems (although extensions to third-order systems are often achieved with the aid of computer graphics). The phenomena of multiple equilibrium points and of limit cycles are clearly seen in phase plane analysis. A number of useful classical theorems for the prediction of limit cycles in second-order systems are also presented.

2.8 Notes and References

Phase plane analysis is a very classical topic which has been addressed by numerous control texts. An extensive treatment can be found in [Graham and McRuer, 1961]. Examples 2.2 and 2.3 are adapted from [Ogata, 1970]. Examples 2.5 and 2.6 and section 2.6 are based on [Hsu and Meyer, 1968].

2.9 Exercises

2.1 Draw the phase portrait and discuss the properties of the linear, unity feedback control system of open-loop transfer function

$$G(p) = \frac{10}{p\,(\,1 + 0.1\,p)}$$

2.2 Draw the phase portraits of the following systems, using isoclines

(*a*) $\ddot{\theta} + \dot{\theta} + 0.5\,\theta = 0$

(*b*) $\ddot{\theta} + \dot{\theta} + 0.5\,\theta = 1$

(*c*) $\ddot{\theta} + \dot{\theta}^2 + 0.5\,\theta = 0$

2.3 Consider the nonlinear system

$$\dot{x} = y + x\,(x^2 + y^2 - 1)\,\sin\frac{1}{x^2 + y^2 - 1}$$

$$\dot{y} = -x + y\,(x^2 + y^2 - 1)\,\sin\frac{1}{x^2 + y^2 - 1}$$

Without solving the above equations explicitly, show that the system has infinite number of limit cycles. Determine the stability of these limit cycles. (*Hint*: Use polar coordinates.)

2.4 The system shown in Figure 2.10 represents a satellite control system with rate feedback provided by a gyroscope. Draw the phase portrait of the system, and determine the system's stability.

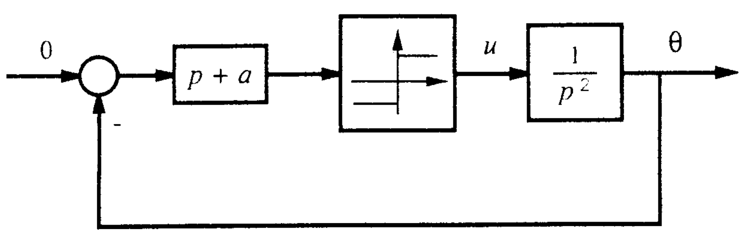

Figure 2.10 : Satellite control system with rate feedback